

Access DB# 224/21

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Amy Nelson Examiner #: 73840 Date: 8/22/00
 Art Unit: 1638 Phone Number 306-418 Serial Number: 08/984 099
 Mail Box and Bldg/Room Location: 9F17 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): McBride et al

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

PLEASE SEARCH SEQ ID NO: 1, 7, 11, 12, 15

PLEASE RUSH!

FOR OFFICIAL USE ONLY

STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher: 224-22 NA Sequence (#) 5 STN 5
 Searcher Phone #: 308-4242 AA Sequence (#) _____ Dialog _____
 Searcher Location: CM112618 Structure (#) _____ Questel/Orbit _____
 Date Searcher Picked Up: 9/1 Bibliographic _____ Dr. Link _____
 Date Completed: 9/5 Litigation _____ Lexis/Nexis _____
 Searcher Prep & Review Time: 4 Fulltext _____ Sequence Systems CompuLink
 Clerical Prep Time: 3 Patent Family _____ WWW/Internet _____
 Online Time: 6 Other _____ Other (specify) _____

Qy 3161 cacacacacaaaaaaaacaaatgtgtgtgtggaattttattattacggaatgtaatat 3200
| | | | | | | | | | | | | | | | | | | | | |
Db 17097 TTATATAATAAATTTATGAATAGGGGGAATAAATTTATTTTCATTTTACATATATATATA 17038

Qy 3221 tatattttaaaaataaaattatgttatttagattcctaataattttggagcattccatacta 3280
| | | | | | | | | | | | | | | | | | | | | |
Db 17037 TATATATATATATACAATTAATTAATTCAGATTAGTGATTAAAAATAAATTATTTTATTA 16978

Qy 3281 taatttcgtacataataattaaaaatagta-atataaagtgtaattaacctttaattac 3339
| | | | | | | | | | | | | | | | | | | | | |
Db 16977 TACTTATATAATTTAATGAAATTAATTTATATGTATATATATATAAATATATGAATTG 16918

Qy 3340 aagcataataattaattttgtaatcaattaattttttattctattattttaatttaattag 3399
| | | | | | | | | | | | | | | | | | | | | |
Db 16917 AATTTTATATAAAATCATTTTAAATTTTATTATATATAAAATATTTTATTAATTTAT 16858

Qy 3400 tctattttttcaaaaataaaatttaaactcaataaaaaataatttttccctaattgtgaaa 3459
| | | | | | | | | | | | | | | | | | | | | |
Db 16857 TAAAAATAATTTTATTATATAAAATAATTTATTATATAAAATAGTTTATTAAGTATAAATTAA 16798

Qy 3460 caactcatgttatacttcaaaattataagttattatttacctgatgatttatttatta 3519
| | | | | | | | | | | | | | | | | | | | | |
Db 16797 TAAATCATTTTTTTTTTAAAAAAAATATTTTTTAAGTTTAAATATACAATAAATTTAT 16738

Qy 3520 gtatattaattctgattataattatggtgggatacaatcgctttccactaaatatttaa 3579
| | | | | | | | | | | | | | | | | | | | | |
Db 16737 GAATAGGGGGAATAAATTTATTTTCATTTTTTATATATATATATATATATA-ATTAA 16679

Qy 3580 ctatgatttataaatttatttcaacatcgatatatttacttattaatacataatttatcat 3639
| | | | | | | | | | | | | | | | | | | | | |
Db 16678 TTATTTTCAGATTTTAGTGATTAATAAATAATTTATTATTATATTTATATAATTTAATGAA 16619

Qy 3640 aattttatggaaattgagaccaagaacatttaagagacaacattctataacaagaacaa 3699
| | | | | | | | | | | | | | | | | | | | | |
Db 16618 AATTAATAATTTATATATATATATATATAAAA-----TATAAATGAAT 16574

Qy 3700 ttagaaaaaatgtactcttttaggtaattttaagtactcttaaccaaacacaaaaattcaa 3759
| | | | | | | | | | | | | | | | | | | | | |
Db 16573 TTTTAAAAAATTTATTTAATTTTATTATAATAAAAAATTTCTTATTAATTTATTTAA 16514

Qy 3760 atcaaatgaactaataagataataataacatacggacatcttacttgaatcttacatt 3819
| | | | | | | | | | | | | | | | | | | | | |
Db 16513 ATAATTTTATTATATAAATAATTTATTATAAAAAAGTTTATTAAGTATAAATTTAATA-- 16454

Qy 3820 cccataattttattatgaaaaataacttatattactcgaactaaatgtgtgcacaaatt 3879
| | | | | | | | | | | | | | | | | | | | | |
Db 16455 --AATAATTTTTTTTTTAAAAAAAATATTTTTTT----AAGTTTAAATATATAATAAA 16403

Qy 3880 attatctaataaagaaaaacacttaatttttataacattttttcatatatttgaagat 3939
| | | | | | | | | | | | | | | | | | | | | |
Db 16402 TTTTGAATAGGGGGAATAAATTTATTTTCATTTTACATATATATATATATATATATA 16343

Qy 3940 tatattttgtatatttacgtaaaaatatttgacatagattgagcaccttcttaacataat 3999
| | | | | | | | | | | | | | | | | | | | | |
Db 16342 TACAATTAATTAATTCAGATTTAGTGATTAATAAATAATTTATTATTACTTATATAAT 16283

Qy 4000 cccaccataagtcagtgatgtagatgagaattggtacaaacacgtggggccaaattccc 4059
| | | | | | | | | | | | | | | | | | | | | |
Db 16282 TTAATGAAAAATTAATTTATGTATATATATATAAATATATGAATTGAATTTTATAAA 16223

Qy 4060 accaaaccatctctcattctctctctataaaaggcttgc--tacacatagacaacaatcca 4117
| | | | | | | | | | | | | | | | | | | | | |
Db 16222 AAATCATTTTAAATTTTATTATATATAAAAAATTTTTTATTAATTTTAAAAATATTTT 16163

Qy 4118 cacacaaatacacgcttcttttctctatttgattaaacctggtctatagcatcgtcac 4177
| | | | | | | | | | | | | | | | | | | | | |
Db 16162 ATTTATAAATAATTTTATTATAAAAAATAGTTTATTAAAGTATAAATTTAATAAATCATTTT 16103

Qy 4178 cettcttctcttttcccaacttttactcataagtgctctactagt----gaccggtagcc 4232
| | | | | | | | | | | | | | | | | | | | | |
Db 16102 TTTTAAAAAAAATATTTTTTAAAGTTTAAATATACAATAAATTTATGAATAGGGGGAA 16043

```

RESULT 15
DMU37541/c
LOCUS      DMU37541      19517 bp      DNA      circular      INV      04-APR-2000
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION  U37541
VERSION    U37541.1      GI:1166529

```

KEYWORDS
SOURCE Drosophila melanogaster.
ORGANISM Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 12511 to 12682)
AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and
Wolstenholme,D.R.
TITLE Drosophila mitochondrial DNA: a novel gene order
JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE 83090428
REFERENCE 2 (bases 5269 to 5695)
AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
flanking sequences and comparisons to mammalian mitochondrial tRNA
genes
JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE 83220794
REFERENCE 3 (bases 404 to 5272)
AUTHORS de Bruijn,M.H.
TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and
genetic code
JOURNAL Nature 304 (5923), 234-241 (1983)
MEDLINE 83245048
REFERENCE 4 (bases 804 to 1778)
AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.
TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in
Drosophila melanogaster and its sibling species
JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE 88174373
REFERENCE 5 (bases 5268 to 13619)
AUTHORS Garesse,R.
TITLE Drosophila melanogaster mitochondrial DNA: gene organization and
evolutionary considerations
JOURNAL Genetics 118 (4), 649-663 (1988)
MEDLINE 88212147
REFERENCE 6 (bases 441 to 2967)
AUTHORS Satta,Y. and Takahata,N.
TITLE Evolution of Drosophila mitochondrial DNA and the history of the
melanogaster subgroup
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE 91088557
REFERENCE 7 (bases 14215 to 14512)
AUTHORS Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and
Atkinson,P.W.
TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are
modified arthropods
JOURNAL Science 258 (5086), 1345-1348 (1992)
MEDLINE 93088057
REFERENCE 8 (bases 14917 to 19517)
AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE Sequence, organization, and evolution of the A+T region of
Drosophila melanogaster mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE 94285822
REFERENCE 9 (bases 1 to 408; 13319 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Drosophila melanogaster mitochondrial DNA: completion of the
nucleotide sequence and evolutionary comparisons
JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE 96423163
REFERENCE 10 (bases 1 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES
source Location/Qualifiers
1. 19517
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/db_xref="taxon:7227"
/note="derived from new and previously submitted

PFMAL1P3/c
LOCUS PFMAL1P3 67970 bp DNA INV 15-DEC-1999
DEFINITION Plasmodium falciparum MAL1P3, complete sequence.
ACCESSION AL031746
VERSION AL031746.9 GI:6594243
KEYWORDS HTG.
SOURCE malaria parasite *P. falciparum*.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 67970)
AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M. and Barrell, B.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT On Dec 16, 1999 this sequence version replaced gi:5763807.
For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from *E. coli*, yeast, vector, page etc.
FEATURES
source Location/Qualifiers
1..67970
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="1"
gene complement(1748..3276)
/gene="MAL1P3.01"
CDS complement(join(1748..2598,2748..2848,2990..3276))
/gene="MAL1P3.01"
/note="MAL1P3.01, conserved hypothetical protein, len: 412 aa, similarity: UPF0006 family eg to YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa overlap)"
/codon_start=1
/product="conserved hypothetical protein, UPF0006 family"
/protein_id="CAB63556.1"
/db_xref="GI:6594244"
/translation="MKLVFHYIKYINLVFYSIIPLKSNLSKIYNDLRYISTVKNKYV LQIKRSNLKKNHNIRKMDNESSPIDIGSNLTKMFDGVYNSKKHENDLQNVNRAK NNNVDKIIITCTCLAEIDKSLKICETYPDEGKFLYSAGVHPTNCEYFDXNKHEEKE IIAKKEYEPIKYFKNQYVENSXKMGKIKCDGKDMNLLNEILLENKNDITPGPKY NEKDKEYLENLKNKIYKPNVICIGELDFDRLYFCSKYIQIKYPIFQLKLQMFN LPMFLHMRNCSETFFKIVDIYKFLFEKNGGVHSFTDKEDIVHIVQYKNLYIGVNG CSLKSLNINAVKKIPLNLLLETDAPWCGVKKTHASYEYIKDYEKRAYTNLKKIKN IIKCDDNTIPKERNPEYNTA"
misc_feature complement(2599..2610)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"
misc_feature complement(2742..2747)
/gene="MAL1P3.01"
/note="potential splice donor sequence, atg/gtaaaa"
misc_feature complement(2849..2861)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"
misc_feature complement(2984..2989)
/gene="MAL1P3.01"
/note="potential splice donor sequence, aaa/gtaaaa"
gene 5005..5496
/gene="MAL1P3.02"
CDS 5005..5496
/gene="MAL1P3.02"
/note="MAL1P3.02, hypothetical protein, len: 163 aa, contains possible signal sequence"
/codon_start=1
/product="hypothetical protein, MAL1P3.02"
/protein_id="CAB63557.1"

/db_xref="GI:6594245"
/translation="MKLLNRRFVLCPIILIFFFLNSVVLGNNRRNNINFHETENAAK AMRKLKSGEINSIKLDNGELKIKLNDKHKDSTKWOKSYFISNLEEEKYSQDLFR KKQEIENEAATKIIEDRQEFYILNNDENIATRFVLENNFDELYIQSFQKSLIDIIQS LNN"
misc_feature 8020..10389
/note="possible cen1, region of very high [A+T] content"
gene 14884..20352
/gene="MAL1P3.03"
CDS 14884..20352
/gene="MAL1P3.03"
/note="MAL1P3.03, putative ABC transporter, len: 1822 aa"
/codon_start=1
/product="putative ABC transporter"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/translation="MTTYKENVGISNKGKNNKSKQNISFLNLSFDWIRPLINDLIK GDIQELPNICRNFDPVYASKLEENLRDIEVDESEFYSEKNSNEHVLHHCNSNDASE KKVNVVYHNILNSILTKFKFRILIIISFYILETLIVTLCGGKIDYYMRILEGQKIPV YISPLKDFKVSGLVVMIMFFHLFFELHFFHLPTINLKSVMYFLYKINLCSNN NHLQNPDAFYNTYKFSQTEIDEISRDLSIGKNASSSSSGIKNNKNNKNNKFFVEN DYIINFIKSTKMEKDLNENRSLPNVNIYIMFSDVPSVTFVTCINLNFVVKIF MSFYVPHIKIGSNSVGIWLSIALYSAMILFEFLPSLFSKYLIYDRKIDNMHVL KEFKLIKMFNWFESPAFKYINIRFMKEMKYKIRLYLSNIGVFISSISDIVEVFIPI YLKDRLNKKKEIKFTSIIMPLVYKILISNVANFPLNVNVMGIVNIKRLNYYINDH LYYNDIKNYFMRYRINEDYIVVDKTFLOENITSHDDGTSHNLKHLKNVKNKLTN MFKYFFPYHKMNYHKNIINKQLSGLLKNVDNNTKKICFQEHKSNTYNYNSSHIHE KKEEYENIHNSNSTMSNEFEKKNKNEYIILENCESFGLSDNKCNDHILKNINFN LKRNLSAIIIGNVSGKSAPFHSILGDFNMTGNYIENFFKMPILYVPQNSWLFMG NIRSMILFGNEYNLIYKYTILOSELNLDSTIEHGMKYINDHNLKSGQKVRICLA RALYEHYIHMHLCTDYEKRLIQNEILDKDLNKNKNISSYNNKSKLVNINIPFEN YLQKCLMDNNFYLYLLDDIFTSIDPSISKRIFSNLFCEDNISFKDNCFSIISMNKS TLDNFLIEDILDNVQYEVNIFEIQDRTKLYRGNISEYMEKNLNITKESHWGYSNLNT IDYTRIKLPDEVELNHHVKSXNMYKEAYFVKGNTESVSFEIDSINKEYIKMKKKNY KKEHMKNKNDNNNNNNNNKDDHINIMNDNHRNYNDINLGNSTDDSPVSSLGNE YTLDTYTSNNSDKEIEVKPLYKDTHEEFNKSSSPFVKSSSNMNNPNSNRYEDNSSI FKGSISLETYLVWYQVQGFVLTSTVIMFLLITTYIIMHPLYASGLIKLYKEKNYVRI LSTLGVCISFSARLGVIIKFLCDYTHIEKEMCCVQRLFEFAKISNKENASMNKENEL NVITTQTYKEKNENISDKISAIVEYKNVLSLSIIINSQDDSEKKKYGIKFENVYVSK KKIPLVNGTYKYIDEEPSLNKNIMYALKNQKIGIVGSGAGKSTILLSILGLINISQK KITVEGRDIRTYNRKGEDSITIGILAQSSSFVYNNWRTFIDPYNNTDDEIVHALKLN GNLGKNDLYKMHKQDMKSNYKIKQTSKVYNQSNNDNTILLTNDICRYLSLVRLYN RHKYKIILDEIPINLNSVHDELSFLIGAKASFNYIIRNHPNNTVLIISHHANT LSCDDIYVLRKGETIYRCSYEDVKTQSELSHLEMD"
rRNA 23896..31533
/gene="rRNA"
/note="region containing small subunit, 5.8S and large subunit rRNA genes and spacer regions"
gene 23896..31533
/gene="rRNA"
gene complement(31966..32775)
/gene="MAL1P3.04"
CDS complement(join(31966..32476,32675..32775))
/gene="MAL1P3.04"
/note="MAL1P3.04, conserved hypothetical membrane protein, len: 203 aa, similarity: P. falciparum chromosome 2, PFB0110W, O96126 predicted integral membrane protein (255 aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity in 191 aa overlap)"
/codon_start=1
/product="conserved hypothetical membrane protein, MAL1P3.04"
/protein_id="CAB63559.1"
/db_xref="GI:6594247"
/translation="MKKSYTFINVTILLFLTLFLTYNYDTFSKTKFNNKIDIN RPKRIIAEASEQKYPEWEDFCLILNEELIRPEHNDSPYLPHEYNIDKINELSINS TKIWKETIKMRQNYEKETDMMNNNRDFFMWHYKMANIYLYKVHKLINITLKDLPNI HDKEETITTWIKIQEDIEYFLNLQVEWLRILTLELFYKNE"
misc_feature complement(32477..32486)

Qy 2219 taactctcaaaatctggtcataacttctaggtgagtttgctgtgctacagttagtaagtc 2278
| | | | | | | | | | | | | | | |
Db 139558 NNNATATATATTTTAAATAATATCATCAA--TTTAATTATAGTTATATAAAGATAAATA 139501

[illegible]

[illegible]

```

RESULT 13
AC008206/c
LOCUS      AC008206   161891 bp    DNA             HTG             08-MAR-2000
Drosophila melanogaster chromosome 3 clone BACR03115 (D765) RPCI-98
03.1.15 map 96B-96B strain y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 133 unordered pieces.
ACCESSION  AC008206
VERSION     AC008206.9   GI:7208834
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 161891)
AUTHORS     Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
             Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
             Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
             Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,

```

Qy 1802 aaataattattaattaaaaatttatggacttttggactgtctgactaattttcagaatttt 1861

Db 75037 AATAAATCTTAATAAATAATTTTTTTGTAGATATTTCTAGGATAATATGAATATTTTC 74978

Qy 1862 attttggttttgggttttggtaattttttagataattttttaattcttcgataatt 1921

Db 74977 GTAAAAAATAATAATAAAAAATGATATTTAAATATTTATTAACATA--CAATATTAAT 74920

Qy 1922 tttctgtattttgaaaggatgttcgaatttttttcaaaattgaaacgtttaagaattt 1981

Db 74919 ATTTTATATTATAAATATTATTATTTAAAAATATATAAATATAAATTATGGAATTTT 74860

Qy 1982 ttactactgcaaaattcagaataagtgaaattgttttttagaagattaanaagtagta 2041

Db 74859 AAATTAATTAATTAATTTGTTTAT-ATTTAATTATATATTTAATAAATATATTTTAAA 74801

Qy 2042 ttacgatttttagtttgggttggaagtaagtgtgtttttgaacataattatttga 2101

Db 74800 ATAATACACACAAAATGATTCTTAATTAATATAAAATATTTATTTATTATATAATA 74741

Qy 2102 caataattaagttttctagggataaacggaaatatcttctctttttgtaaaattact 2161

Db 74740 TAAATAATTTAATTTAAATAAAATAAAATAAAATAAATAATTATTATTATATAATA 74681

Qy 2162 aatgcagaacaacaacgcttttggggagcaataatctagctttaaagtagtcagttaa 2221

Db 74680 TATTTAATTTAATTTAATTTATTATTATTAAATTTAAATAAAATAAAATAAATAT 74621

Qy 2222 ctctcaaaatctggtcataactcttaggctgagtttgcgtgtcacagtagtaagctat 2281

Db 74620 TTATTATTATATAATATATTTTAATTAATTAATTTATTATTAAATTAATTAATTTAAA 74561

Qy 2282 agaaacttacctgcacaaacgacatgacgtcagggtcgaaatctacaacttttccctttc 2341

Db 74560 TAAAAATAAAATAAATATTTATTATTATATAATATATTTAATTAATTAATTTATTA 74501

Qy 2342 ttcaattaacatagtggttgattcaagttccgatctataataattattacagttatcaa 2401

Db 74500 TTTAATTAATTTAAATAAATAAATAATTAATTAATTAATTTAATTAATTAATTAATA 74441

Qy 2402 ttcaattaccttatatcatcctattataaataaagtcagttcaattcagtttgcgaa 2461

Db 74440 AAATAATATCAATTATAAATTTATTAAATTTTAAATAAATAAATTAATAATAATAA 74381

Qy 2462 gtcccaaaaaattttgaaattttataaaattttccctaaaaccgaaatagttatattct 2521

Db 74380 ATTAATAATTTAATTAATAATAAATAACATTATAAATTAATAAATAATTTAATAAAT 74321

Qy 2522 tcaaaatttaagtttcatttttcaatccgatttcaatttcatctctttataactctctatt 2581

Db 74320 ATATATAATTATTATAAATTTAAATACATATATTTTAAATAAATTTCTTAATTTAT 74261

Qy 2582 atctataattacataaaatttcaaatattttgaaattatttcaacttttagtccctaaagt 2641

Db 74260 TTATTACATTATTATAATATATTTTTTATTAAAAAATAAATAAATAAATTAAT 74201

Qy 2642 caaaactataaatttttcaacttttagaanaattatcatttttccacatctaagcatcaaat 2701

Db 74200 TAATAATTAATATATTTAATAGTAATTAATATTAACAAATAAATAAATTAATATA 74141

Qy 2702 accaaatgacacaaatttcatgattagttagatcaagcttttgagttctcaaacataaa 2761

Db 74140 ATATTATTAATTAATTAATAATAAATTTATTATTATAATATATATATATTAATAAT 74081

Qy 2762 aattacaaaaaaacaaactttaaataattttatcatcaattttgaacaacaagcttggcgg 2821

Db 74080 TAATTTAAATATAAATTAAGAAGATAATTTTATACTTTATTATTAATTAATATATAGT 74021

Qy 2822 aatgctaagagcttaaaaaatggctcttttggtttctttttgttgcacagctggagagaa 2881

Db 74020 AATAAATAATTTATGTTATTATTATAAATAATTTATTATTATTATTATTATTATTAA 73961

Qy 2882 gaggygaatgaagattgaccatattttttattatgttttaacat---ataatattaat 2937

Db 73960 TAAATTAATTAATTTTAAATAATATTTATTATTTAAATTAATAATAACATATAATTA 73901

JOURNAL Science 258 (5086), 1345-1348 (1992)
MEDLINE 93088057
REFERENCE 8 (bases 14917 to 19517)
AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE Sequence, organization, and evolution of the A+T region of
Drosophila melanogaster mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE 94285822
REFERENCE 9 (bases 1 to 408; 13319 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Drosophila melanogaster mitochondrial DNA: completion of the
nucleotide sequence and evolutionary comparisons
JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE 96423163
REFERENCE 10 (bases 1 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES
source Location/Qualifiers
1. .19517
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/db_xref="taxon:7227"
/note="derived from new and previously submitted
sequences; sequence is a composite containing sequences
obtained from different Drosophila melanogaster strains"
trNA 1. .65
/gene="mt:ND6"
/product="tRNA-Ile"
/db_xref="FlyBase:FBgn0013685"
gene 1. .19517
/gene="mt:ND6"
/note="mitochondrial NADH-ubiquinone oxidoreductase chain
6"
/allele=""
/db_xref="FlyBase:FBgn0013685"
trNA complement(97. .165)
/product="tRNA-Gln"
trNA 171. .239
/gene="mt:ND6"
/product="tRNA-Phe"
/db_xref="FlyBase:FBgn0013685"
CDS 240. .1265
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC47811.1"
/db_xref="GI:1166530"
/translation="MFNNSKILFITIMIIGTLITVTSNWLGAWMGLEINLLSFIPL
LSDNNMLMSTEASLKYFLTQVLASTVLLFSSILLMLKNNMNEINESFTSMIIMSALL
LKSAAAPFHFPPNMMEGLTMMNALMLMTWQKIAPLMLISYLNKYLILISVILSVII
GAIGGLNQTSLRKLMAFSSINHGLWMLSSLMISESIWLILFFPYFSLFVLTFFMNFIF
RLPHLNQLPSWVNSKILKFTLFMNFSLGGLPPFLGFLPKWLVIOQLTLCNQYFMLT
IMMSTLITLFFYLRYCYSAPMNYFENNIMKMMNSINTNMYMINTFFSIFGLFLI
SLFYFMF"
trNA 1264. .1329
/gene="mt:ND6"
/product="tRNA-Trp"
/db_xref="FlyBase:FBgn0013685"
trNA complement(1322. .1383)
/product="tRNA-Cys"
trNA complement(1403. .1468)
/product="tRNA-Tyr"
CDS join(1470. .1472,1474. .3009)
/codon_start=1
/exception="mechanism underlying reading frame shift after
first codon uncertain"
/transl_table=5
/product="cytochrome c oxidase subunit I"
/protein_id="AAC47812.2"

/db_xref="GI:7412849"
/translation="MSRQWLFSTNHRDITGLYIFGAWAGMVGTSLILIRAEGLHPG
ALIGDDQIYNIVTAAHAFIMIFFMVPIMIGGFGNWLVPMLGAPDMAPPMNNMSFW
LLPPLSLLLVSSWVENGAGTGTWVYPPSLAGIAGGASVDLAIFSLHLAGISSILGA
VNFITTVINMRSTGISLDRMPLFVSVVITALLLLSLPVLAGAITMLLTDRLNLSF
FDPAGGGDPILYQHLFWFFGHPEVYILILPGFGMISHIISQESCKKETFGSLGMIYAM
LAIGLLGFIWVAHMTFVGMVDVTRATYTSATMIIAVPTGKIFSWLATLHGTQLSYS
PAILWALGFVLEFTVGLTGVLVANSVDIILHDTIYVVAHFHYVLSMGAVFAIMAGF
IHWYPLFTGLTLMNKLKSHFIIMFIGNVLTFFPQHFLGLAGMPRRYSYDAYTTWN
IVSTIGSTISLLGILFFFFIIESLSVSRQVIYPIQLNSSIEWYQNTPPAHSYSELP
LLTN"
trNA 3012. .3077
/gene="mt:ND6"
/product="tRNA-Leu"
/db_xref="FlyBase:FBgn0013685"
3083. .3767
/note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/codon_start=1
/transl_except="(pos:3767,aa:TERM)
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="GI:1166532"
/translation="MSTWANLGLQDSASPLMBQLIFFHDHALLILVMITVLVGLMFM
LFFNNYVNRFLHGLLEMIWITLPAIILLFIALPSRLRLVLLDEINEPSVTLKSGH
QWYWSYETSDFNIEFDSYMIPTELMDGFRLLLDVDRVLPMSQIRILVTAADVI
HSWTVPALGVKVDGTPGRNLNINFFINRPLGYQCSEICGANHSFMPVIESVPVNY
FIKWISSNNS"
trNA 3768. .3838
/gene="mt:ND6"
/product="tRNA-Lys"
/db_xref="FlyBase:FBgn0013685"
trNA 3840. .3906
/gene="mt:ND6"
/product="tRNA-Asp"
/db_xref="FlyBase:FBgn0013685"
CDS 3907. .4068
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="ATPase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/translation="MPQMAPISWLLFIIFSITFLFCSINYSYMPNPKSNELKNI
NLNSMNMWKN"
CDS 4062. .4736
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="ATPase 6"
/protein_id="AAC47815.1"
/db_xref="GI:1166534"
/translation="MTNLFVSVDPLAIFNFWSLNWLSTFLGLLMIPSIYWLMPSPRYNI
MWSILLTLHKEFKTLGSPSGHNGSTFFISLFLFNNFMGLFPYIFTSTSHLTAT
LSLALPLWLCFMYGWINTHQMFARLVPQGTAILMPFVCITETISNIRPGTLAVR
LTANMIAGHLLTLTGNTGSSSYMLMTFLMAQIALLVLESAMIQSYVFAVLSTL
YSSEVN"
CDS 4736. .5524
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AAC47816.1"
/db_xref="GI:1166535"
/translation="MSTHNSPHFLVDYSPWPLTGAIGAMTTVSGMVKVHQYDISLF
VLGNIITLTVQWNRVDSREGTYQGLHTYAVTIGLRWGMILFILSEVLFFVSFFWAF
FHSSLSPAIELGASWPPMGIISFNPQIPQLLNTALLASGVTVTVAHSLMENNHSQT
TQGLFTVLLGIYFTIQAAYEIEAPFTIADSIYGSTFFMATGFHGIHVLIGTTFLLV
CLLRHLNNHFSKNHFGFEAAAWYHWFVDVWVFLYITITYWGG"

Query Match 3.4%; Score 189.6; DB 33; Length 4601;
Best Local Similarity 44.3%; Pred. No. 3.7e-13;
Matches 1063; Conservative 0; Mismatches 1309; Indels 28; Gaps 6;

Qy 1538 ttttcgaaaaaaatttgcattgtgtttttctgaaaaaattgcattacaataatcatgca 1597
||| ||||| | | | | | | | | | | | | | | | | | |
Db 2203 TTTTAAAAAAATTTATTATAAATTATACTTAATAAACTATTTTATAATAAATTA 2262

Qy 1598 ttctcaattttgtgccaattgaacgttataaaaattctctatgatctgatctgtttatt 1657
| | | | | | | | | | | | | | | | | | | | | | |
Db 2263 TTTTATAATAAAATTTATTTAAAAATAATAAAAAATTATATATATATATATATATA 2322

Qy 1658 acattatatg-tgtttatgcttgagtttaagtcacaacattgagattcatagctcacccaat 1716
| | | | | | | | | | | | | | | | | | | | | | |
Db 2323 TATTAATGAAAAATATTTTAAATTTTAATAATAAATAAATTAATAATTAATAATTA 2382

Qy 1717 tatttaacatttcaggcaatctgcagacttaggattggatggcgttcaggagcttggat 1776
| | | | | | | | | | | | | | | | | | | | | | |
Db 2383 AATAAAATCTATTCATTATTAATATTTAATTAATAATAAATAAATTAATAACTAATAAT 2442

Qy 1777 tggttttctcacatcatatttttataaataattattaataaaattatggacttttga 1836
| | | | | | | | | | | | | | | | | | | | | | |
Db 2443 TAAATAAAATTTATTTACTAATATTTAATTAATAATAAAAAATTTATTTTTTTTTT 2502

Qy 1837 ctgctg-actaatttccagaattttatttgggtttgggttttggtaatttttaga 1894
| | | | | | | | | | | | | | | | | | | | | | |
Db 2503 TTTTTTTTAAATAATTAATTAATTTATATATTTTATAAATTTATATATTATTGAATAT 2562

Qy 1895 taattattttaaatattctgcataatttttctgtatttggaaaggatgttcgaatttt 1954
| | | | | | | | | | | | | | | | | | | | | | |
Db 2563 TTATAATATATATATATATAGAAAAATTAATTTATTAATAATTTAATATAAATTTT 2622

Qy 1955 tttcaaaattgaaacgtttaagaatttttactactgcaaattccagaataagtgaattgt 2014
| | | | | | | | | | | | | | | | | | | | | | |
Db 2623 TTAATAAATTCCTAAATGTAATTTTATAAAAAATTTATATATAAATAATCATGTTT 2682

Qy 2015 tttttgaaagattaaataa--gttagtattacgatttttagtttgatttgggtgaaagt 2072
| | | | | | | | | | | | | | | | | | | | | | |
Db 2683 TTTAAAAATAAACAAAAAATTTTAAATAAATAAATTTATAATGAATATAATTTATTT 2742

Qy 2073 aatgtatgtttttgaacataattatttgacaataaataagttttctagggataaacgga 2132
| | | | | | | | | | | | | | | | | | | | | | |
Db 2743 ATTTTCATTTTAAAAAAATTTTAAAAAAATAATTTTTTTTAAAAAAAACT 2802

Qy 2133 aatatctctctctttttgtaaaattactaatgcaagaacaacaacgttttggggagca 2192
| | | | | | | | | | | | | | | | | | | | | | |
Db 2803 ATATACTAATTATAAATTAATAGATATTATATATATATAAATTTAATATATTATTAT 2862

Qy 2193 aataatctagctttaagtagtcagtgtaactctcaaatctggtcataactcttaggctg 2252
| | | | | | | | | | | | | | | | | | | | | | |
Db 2863 ATATCTAATAATTTAAATAAAAAATTTTAAATTTAAATGATAGATATAATTTATAAAA 2922

Qy 2253 agtttctgtgctacagtagtaagctctatagaacttacctgacaaaacgacatgacgtc 2312
| | | | | | | | | | | | | | | | | | | | | | |
Db 2923 ATTTATATTCATATTTAATTTATTATTAATTTAATTTATATAAATAATATAAATTTA 2982

Qy 2313 agggtcgaatctacaacttttctttttcttcaattaacatagtttattcagttccg 2372
| | | | | | | | | | | | | | | | | | | | | | |
Db 2983 ATTAATTATTATATATTTATAAATTTATATATTATTGAATTTATATAATATATATATA 3042

Qy 2373 atctataataatttattacgatttatcaatttcaattaccattatcatcattataaaa 2432
| | | | | | | | | | | | | | | | | | | | | | |
Db 3043 TATATAGAAAAATTAATTTATTAATAATTTAATAAATTTTAAAAATTTCTTAAA 3102

Qy 2433 tataagtcagttcaattcagttttcgaaagttcccaaaaatttgaattttataaattt 2492
| | | | | | | | | | | | | | | | | | | | | | |
Db 3103 TGTATTATTTTATAAAAAATATTTATATAATAAATCATGTTTTTAAAAAATAACAA 3162

Qy 2493 attccctaaaaccgaatagtatatctttcaaatttaagtttcatttttcaatccgatt 2552
| | | | | | | | | | | | | | | | | | | | | | |
Db 3163 AAAATTTTAAATAAATAATTTTATAATGAAATAAATTTATTATTTTCAATTTTTT 3222

Qy 2553 tcaatttcatccttttataactctctattatctataattaccataaaatttcaaaattt 2612

```

/gene="MAL3P5.7"
/note="predicted using hexExon; MAL3P5.7 (PFC0605c),
Hypothetical protein, len: 248 aa"
/codon_start=1
/protein_id="CAB41709.1"
/db_xref="GI:4725991"
/db_xref="SPTREMBL:Q9Y011"
/translation="MGHGGLNLPLQKKWNVYRKDAQYKVHYDEHKIIKEEKDEIKR
KKDEFESTISTLKKNTKNEEDNNYNNFYDENGKKTNTNYCNDHINLFIDEEKELT
AKQKKHEEFLIKKGHYIYDKNFNTQHNSIYDKNKNAQIISDFNMKLCERDWFNLKK
NKNKTKDNGANFPHIQKDNISEENKNTENISDLSLYCNTNNYITHDKKKEKKQMY
HIKKIIKYQEKDEKKRKRQGEKKKPK"
gene complement(29992..33537)
/gene="MAL3P5.8"
CDS complement(29992..33537)
/gene="MAL3P5.8"
/note="predicted using hexExon; MAL3P5.8 (PFC0610c),
Hypothetical protein, len: 1182 aa"
/codon_start=1
/protein_id="CAB38971.1"
/db_xref="GI:4493935"
/db_xref="SPTREMBL:O97260"
/translation="MAHVKKREKTEAQETPVVAKEQTHAKEENNESNIAVTEENVIS
KNGQEIATSKNDQEIATSKNDQEIATSKNDQEIATSKNDQENVALNSSEERQNASKEE
DNLQIKQEPHFDISNENEHNENRSPSTLSSFFKEYEENSVEQHFSEGTTEHSME
DSNNVETIENAITNDVLRNRSRSTYSKQKNELTSVTCVCGETVDLNIWSDHIFAKHL

```

Journal of Management Education 30(6)p.789-804
© The Author(s) 2006. Reprints and permissions:
<http://www.sagepub.com/journalsPermissions.nav>


```

LOCUS      I40338          1283 bp    DNA             PAT      13-MAY-1997
DEFINITION Sequence 17 from patent US 5620882.
ACCESSION  I40338
VERSION    I40338.1  GI:2082630
KEYWORDS
SOURCE     Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE  1 (bases 1 to 1283)
  AUTHORS  John,M.
  TITLE    Genetically engineering cotton plants for altered fiber
  JOURNAL  Patent: US 5620882-A 17 15-APR-1997;
  FEATURES  Location/Qualifiers
            source          1. .1283
                           /organism="unknown".
BASE COUNT      509 a      233 c      251 g      290 t
ORIGIN

```

Query Match 5.0%; Score 273.4; DB 5; Length 1283;
 Best Local Similarity 84.2%; Pred. No. 2.1e-22;
 Matches 326; Conservative 0; Mismatches 46; Indels 15; Gaps 1;

[illegible]

```

RESULT      6
AC005504
LOCUS       AC005504   104992 bp   DNA           HTG           01-APR-1999
DEFINITION  Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC005504
VERSION     AC005504.3   GI:4558584
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1 (bases 1 to 104992)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
            and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 104992)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA

```

94304, USA

COMMENT On Apr 2, 1999 this sequence version replaced gi:4337172.

- * NOTE: This is a 'working draft' sequence. It currently
- * consists of 3 contigs. The true order of the pieces
- * is not known and their order in this sequence record is
- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

```
*      1      58642: contig of 58642 bp in length
*      58643  58842: gap of unknown length
*      58843  91011: contig of 32169 bp in length
*      91012  91211: gap of unknown length
*      91212 104992: contig of 13781 bp in length.
```

```

FEATURES
  source      1. 104992
              /organism="Plasmodium falciparum"
              /db_xref="taxon:5833"
              /chromosome="12"

```

BASE COUNT 44286 a 9326 c 9564 g 41411 t 405 others
ORIGIN

Query Match 3.6%; Score 197; DB 41; Length 104992;
Best Local Similarity 45.7%; Pred. No. 2.2e-14;
Matches 1118; Conservative 0; Mismatches 1290; Indels 37; Gaps 11;

DQ 1556 atttgtgttttctgaaaaattatgcattaacataatcatgcattctcaatttggtcagt 1615
 ||| |||| | | | | | | | | | | | |
DB 72352 ATTTTATATTATTTCATAAAAAAGGATAAAGCAATAATAAAATTATAATAAAAAAAAAACA 72411

Oy 1616 tgaacgttataaaaattctctatgatatacctgatctggtttattaccattatatgtgtttatg 1675
| | | | | | | | | | | | | | | |
Db 72412 TTCCAAAATATACCCCAAAATATATAATTATATAGTATAGAAGACTACACAATATATAAAA 72471

Qy 1676 cttgagttaagtcaaacattgagattcatagctcacccaattatattaatcatttcaggca 1735
| | | | |
Db 72472 TATAAATAATCACATATTAAATATAATATATTTATTATAATTATATAAAAATAATAAAAAAT 72531

QY 1736 atctgcagacttaggattggatggcgttcaggagcttgattggtttctcacatcata 1795
|| || | || | || | || | || |
Db 72532 ATATATAAATAATATATAATATACAAATTAAATCATATAAAAATTATTAAAAATATATT 72591

Qy 1796 tttattaataatattattaataaaaatttatggacttttggactgtctgactaatttcag 1855
||||||| .||| ||||| | | | | |
Db 72592 AAAATTAATAATATATATAATTAAATAAATAATTAAGCTTAAATTATTTAATAAATAAAAAATAA 72651

Qy 1856 aat ttttatttttggtttgggttttgtgaatttttagataatttttaaatattctgc 1915
| | ||| | | | | | | | | | | | |
Db 72652 TAATAAAATTAATATTAAATATAAATAAAATTCATAATACACATTAATTAATAAAATATGA 72711

Qy 1916 ataatttttctgttatTTgaaaaggatgttcgaatttttttcaaaattgaaacgtttaa 1975
||| | | || | | ||| ||| ||| ||| |||
Db 72712 ATATTAAATATAAATAAATAGAAAAATATTTAACAACTTAAATATTAAATAAATAAA 72771

Qy 1976 gaatttttactactgcaaatccagaataagtgaaattgttttttagaagattaaataag 2035
 | | | | | | | | | | | | | | | | | | | | | |
 Db 72772 AATATTATAAATTTATAAATAATAAATATTAATATAAATTAATTAATAATATAATAA 72831

Qy 2036 ttagtattacgattttt---agtttgatttggtgaaagtaatgtatgttttgaacata 2092
||| ||| | | ||| | ||| ||| ||| | ||| |||
Db 72832 TTAATATAATTAAATTTAAATATAAATTAAATAAAAAAATACTAATATTAATATAAA 72891

Qy 2093 attatttgacaataattaagtttcttagggaataaacggaaatatcttctctttttgt 2152
|| | | ||| | | | | | | | | | | |
Db 72892 ATAAAAATATAAATAAAATTATTAATAAAATTAAT--ATAAAAATATAATATAAT 72949

Qy 2153 aaaattactaatgcaagaacaacacgttttggggagcaataatctagctttaagttag 2212
|| | ||| ||| ||| ||| ||| ||| |||
Db 72950 TAATTAATAATAAATAATAAAATTAATTAATAATAATAATAATAAATAAATAAATTAATTAATTAAC 73009

Oy 2213 tcagttaactctcaaaatctggtcataacttctaaggc---tgaqtttgctgtgctacag 2269

```

RESULT 1
I18362
LOCUS      I18362      1283 bp      DNA              PAT      07-OCT-1996
DEFINITION Sequence 17 from patent US 5495070.
ACCESSION  I18362
VERSION    I18362.1  GI:1598717
KEYWORDS   .
SOURCE     Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE  1  (bases 1 to 1283)
AUTHORS   John,M.
TITLE      Genetically engineering cotton plants for altered fiber
JOURNAL    Patent: US 5495070-A 17 27-FEB-1996;
FEATURES   Location/Qualifiers
            source      1. .1283
                       /organism="unknown"
BASE COUNT 509 a      233 c      251 g      290 t

```

```

RESULT 15
AW688369
LOCUS      AW688369      625 bp      mRNA                      EST          17-APR-2000
DEFINITION NP006F02ST1F1000 Developing stem Medicago truncatula cDNA clone
NP006F02ST 5', mRNA sequence.
ACCESSION  AW688369
VERSION    AW688369.1  GI:7563105
KEYWORDS   EST.
SOURCE     barrel medic.
ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Euglenophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
            Fabaceae; Papilionoideae; Medicago.
REFERENCE  1 (bases 1 to 625)
AUTHORS    Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
            Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.
TITLE      Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
            Center for Medicago Genomics Research
JOURNAL    Unpublished (2000)
COMMENT    On May 20, 1999 this sequence version replaced gi:4878271.
            Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA

```

Search completed: September 2, 2000, 22:58:41
Job time: 19337 sec

Db 299 TATGGGACACTGCAGGGCAAGAAGATTACAATAGGCTGAGGCCCTTAAAGCTATAGAGGAG 358
Qy 241 ctgattgtgtttttgttgcccttttcttataagcaagccaggttatgaaacatctaca 300
Db 359 CTGATGTGTTTTTGTGTGCTATTCTCATCAGCAAGCCAGTTATGAGAACATCTCCA 418
Qy 301 aaaagtggatccagagctaagacattatgctcataatgtaccagttgtgcttggtaa 360
Db 419 AAAAGTGGATACCTGAGCTGAGACATTATGCTCCANATGTGCCTATAGTGTGTTGGAA 478
Qy 361 ccaactagatttgcgagatgacaagcagttcctcattgatcacctggagcaacaccaa 420
Db 479 CAAACTAGATNTGCGAGATGANCAGCAATNTCTGATTGATCATCCGGGATCCGCACGAA 538
Qy 421 tatcaacatctcagga 437
Db 539 TAACAACCTGCTCAGGCA 555

RESULT 12

AI900170
LOCUS AI900170 658 bp mRNA EST 06-DEC-1999
DEFINITION sc01g12.y1 Gm-cl012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl012-959 5' similar to SW:RAC1_PEA Q35638 RAC-LIKE GTP BINDING
PROTEIN RH01. ; mRNA sequence.
ACCESSION AI900170
VERSION AI900170.1 GI:5606072
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 658)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT On Oct 30, 1998 this sequence version replaced gi:3812130.
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Ghbc
High quality sequence stop: 401.

FEATURES
source Location/Qualifiers
1. 658
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl012-959"
/clone_lib="Gm-cl012"
/tissue_type="Apical shoot tips, 9-10 day old etiolated
seedlings"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from the apical shoots of 9 to 10 day old etiolated
seedlings. The shoot tips including any emerged leaves
were harvested for mRNA isolation. The cDNA library was

prepared using the Stratagene pBluescript II XR cDNA
library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelding."

BASE COUNT 176 a 125 c 171 g 184 t 2 others
ORIGIN

Query Match 33.6%; Score 305.6; DB 45; Length 658;
Best Local Similarity 73.9%; Pred. No. 7.7e-64;
Matches 386; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 5 aacaatgagcactgcaagatttatcaagtgtgtcacggtcggtgagctgtggggaa 64
Db 136 AAAGATGAGCGCTTCTAGGTTTCATCAAGTCGCTACTGTTGGGGATGGTGTGGGCA 195
Qy 65 aactgtatgtctcatttcataaccagcaactcttcccaacggattatgtccaacagt 124
Db 196 AACCTGCTGCTTATTCTCTACACCGAACACTTTCCCAACCGATTGTGCCAGTGT 255
Qy 125 atttgataactttagtccaatgtggtgtggtgagcagcagtgaaaccttggcctatg 184
Db 256 TTTTGACAATTCAGTCAAATGTGGTGTCAATGGGAGCATTTGTGAATCTGGGTTTGTG 315
Qy 185 ggacactgccgggcaagaagattataataggctaagccactgagttatagaggagctga 244
Db 316 GGATACCTGCTGGACAAGAGGATTACAACAGATTAAAGACCTTTGAGTTACCGTGGTCCGA 375
Qy 245 tgtgtttttgttgcccttttcttataagcaagccaggttatgaaacatctacaaaa 304
Db 376 TGTTTTCATATTGGCTTTCTCTCATAGCAAGGCCAGTTATGAAATGTCTTAAAAA 435
Qy 305 gtggatccagagctaagacattatgctcataatgtaccagttgtgcttggtaacaa 364
Db 436 GTGGATTCCAGAGTTGAAGCATTATGCACCTGGTCCCACTATTCTGGTTGGCAGANA 495
Qy 365 actagatttgcgagatgacaagcagttcctcattgatcacctggagcaacaccaatc 424
Db 496 GCTTGACCTTCGGGATGATTAGCAGTTCTGCATCGACCATTCGTGGTCCGCTACCTATTAC 555
Qy 425 aacatctcaggagaagaactaagaagatgataggagcagttactatatagaatgcag 484
Db 556 CACAGCTCANGGAGAAGAGCTTAGGAAGCTGATTATGACCAGCTTACATTGAATGCAG 615
Qy 485 ctccaaaacccaacagaaatgtgaaggtgttttgcagctgc 526
Db 616 TTCAAAAACACAGGAGAAGCTGGATGCAGCTTTTGATGCAGC 657

RESULT 13

AI901141
LOCUS AI901141 549 bp mRNA EST 06-DEC-1999
DEFINITION sc21b12.y1 Gm-cl013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl013-1272 5' similar to SW:RACD_GOSHI Q41253 RAC-LIKE GTP
BINDING PROTEIN RAC13. ; mRNA sequence.
ACCESSION AI901141
VERSION AI901141.1 GI:5607043
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 549)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,

ACCESSION AI900160
VERSION AI900160.1 GI:5606062
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 688)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R. and Wilson, R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT On Jul 30, 1997 this sequence version replaced gi:2286374.
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 431.

FEATURES
source Location/Qualifiers
1. 688
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1012-936"
/clone_lib="Gm-c1012"
/tissue_type="Apical shoot tips, 9-10 day old etiolated
seedlings"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from the apical shoots of 9 to 10 day old etiolated
seedlings. The shoot tips including any emerged leaves
were harvested for mRNA isolation. The cDNA library was
prepared using the Stratagene pBluescript II XR cDNA
library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelding."

BASE COUNT 191 a 132 c 179 g 184 t 2 others
ORIGIN

Query Match 33.8%; Score 307.2; DB 45; Length 688;
Best Local Similarity 74.1%; Pred. No. 3.2e-64;
Matches 387; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 5 aacaatgagcactgcaagatttatcaagtgtgtcagcggtcggtgagctgtgggaa 64
|| ||||| || || ||||| ||||| || ||||| ||||| ||
Db 166 AAGATGAGCGCTTCTAGGTTTCATCAAGTCGCTACTGTTGGGATGGTCTGGGCAN 225

Qy 65 aactgtgtatgctcatttcataaccagcaactcttcccaacggattatgttccacagt 124
||| || ||||| || ||||| ||||| || ||||| || |||||
Db 226 AACCTGCTTCTTATTCTTACACCAACACATTTCCCAACCGATTATGTGCCGACTGT 285

Qy 125 atttgataacttttagtgccaatgtggtggtggatggcagcacagtgaacctggcctatg 184
||||| || || ||||| ||||| || ||||| ||||| || |||||
Db 286 TTTTGACAAATTCAGTGCAAATGTGGTTGTCAATGGGAGCATTTGGAATCTGGGTTTGTG 345

Qy 185 ggacactgccggggaagaagattataataggcttaagccactgagttatagagagctga 244
||| ||||| || ||||| ||||| || ||||| ||||| || |||||
Db 346 GGATACTGCTGGACAAGAGGATTACAACAGATTAGACCTTTGAGTTACCGTGGTGCCGA 405

Qy 245 tgtgtttttgttggcctttttctttataagcaaggccagcttatgaaacatctacaaaa 304
||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 406 TGTTTTCATATTGGCTTTCTCTCATAAGCAAGCCAGTTATGAAAATGCTCTATAAAA 465

Qy 305 gtggatccagagactaagacattatgtctcataatgtaccagttgtgttgggaaccaa 364
||||| ||||| || ||||| ||||| || ||||| ||||| || |||||
Db 466 GTGGATTCAGAGTTGAAGCATTATGCACCTGGTGCCCATTTATTCTGGTGGCAGAAA 525

Qy 365 actagatttgcgagatgacaagcagttcctcattgatccctggagcaacaccaatc 424
|| || || ||||| ||||| ||||| || ||||| ||||| || |||||
Db 526 GCTTGACCTTCGGGATGATAAGCAGTTCTGCATCGACCATCTGGTGCCGTACCTATTAC 585

Qy 425 aacatctcagggaagaactaagaagatgataggagcagttacttatagatgcag 484
||| ||||| ||||| || ||||| ||||| || ||||| ||||| |||||
Db 586 CACAGCTCANGAGAAGAGCTTAGGAAGCTGATTAATGACCAGCTTACATTGAATGCAG 645

Qy 485 ctccaaaacccaacagaatgtgaaggctgttttccgatgctgc 526
|| ||||| || || ||||| || ||||| || ||||| || |||||
Db 646 TTCAAAAACACCAGGAGAGCTGAAGGCAGTCTTTGATGCAGC 687

RESULT 10

AW705028

LOCUS

AW705028 592 bp mRNA EST 18-APR-2000
DEFINITION sk41f03.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1019-5142 5' similar to SW:RAC5_ARATH Q38937 RAC-LIKE GTP
BINDING PROTEIN ARAC5. [1] ;, mRNA sequence.

ACCESSION AW705028

VERSION AW705028.1 GI:7589250

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 592)

AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V.,
Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J.,
Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M.,
Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N.,
Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT On Jun 22, 1998 this sequence version replaced gi:3246649.

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 430.

FEATURES

source

Location/Qualifiers
1. 592
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-5142"

University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612-625-7219
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name: M250659e ; TIGR sequence name: MTBAD48TK ; More
information, including clone ordering, is available at: .
'http://chrysie.tamu.edu/medicago'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

```

FEATURES             Location/Qualifiers
     source            1..669
                        /organism="Medicago truncatula"
                        /cultivar="genotype Al7"
                        /db_xref="taxon:3880"
                        /clone="pDSIR-7H24"
                        /clone_lib="DSIR"
                        /tissue_type="infected root"
                        /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
                        XhoI; roots infected with Phytophthora medicaginis "
BASE COUNT            175 a    125 c    162 g    205 t     2 others
ORIGIN

```

Query Match 35.6%; Score 324.4; DB 74; Length 669;
Best Local Similarity 75.2%; Pred. No. 2.2e-68;
Matches 403; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 2 aaaacaatgagcactgcaagatttatcaagtgtgtcacggtcgggtgatggagctgtggg 61
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 133 AAAGAAGATGACGCCTTCTAGGTTTCATCAAGTGTTACTGTTGGGGATGGAGCTGTTGG 192
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 62 gaaaacttgtagtctcatttcataataccagaacatactttcccacaggattatgttccaac 121
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 193 TAAAACTGTGTTGTTAATTTCATACACCAGCAATACCTTCCCAC TGACTATGTGCCAAC 252
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 122 agtatttgataacttttagtgccaatgtggtggtggatgcgcacagtgaaacctgtgect 181
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 253 TGCTCTCGACAATTCAGTGCAAATGTGTTGTGAATGGAAGCACTGTGAATCTGGGTTT 312
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 182 atgggacactgccggcggaagattataataggctaaggccactgagttatagaggagc 241
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 313 GTGGGACACTGCAGGACAAAGAGATTATAACAGATTAA GACCTTTGAGTTATCTGGTG C 372
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 242 tgatgtgtttttgttgcccttttctctatatagcaaggccagttatgaaaacatctacaa 301
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 CGATGTTTTCATCTCGCTTTCCTCCCTATAAGCAAGGCCAGTTATGAAAATGTTCCAA 432
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 302 aaagtggatccccagagctaagaacattatgctcataatgtaccagttgtgctgttggaac 361
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 433 AAGCTGGATTCCAGAGTTGAAGCATTATGCACCTGGTGTTCCCATTAATCTGTTGGCAC 492
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 362 caaatagatattgfcgagatgacaagcagttcctcattgatccacctggagcaacaccaat 421
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 493 AAAGCTTGACCTTCGGGATGACAAGCAGTTCTTCGTGACCATCCAAGTGCTGTCTCTAT 552
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 422 atcaacatctcagggagaagaactaaagaagatgataggagcagttactatatagaaatg 481
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 553 TACCACCTGCTCANGAGAGAAGACTTANGAAGCTGTATCAATGCACCTGCTTATATCGAATG 612
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 482 cagctccaaaaaccaacagaatgtgaaggctgttttcgatgtgcataaaagtag 537
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 613 CAGTTCGAAATCACAGCAGATGTGAAAGCAGTCTTTGTATGCACCATTAAGAGTTG 668
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 7
AW349629/c
LOCUS AW349629 796 bp mRNA EST 01-FEB-2000
DEFINITION GM210005B21A12R Gm-r1021 Glycine max cDNA clone Gm-r1021-1584 3',
mRNA sequence.
ACCESSION AW349629
VERSION AW349629.1 GI:6847339
KEYWORDS EST.
SOURCE soybean.

```

ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales;
Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 796)

AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
Erpelding,J., Raph,C., Shoop,E., Pardinass,J., Liu,L. and Lewin,H.

TITLE A Functional Genomics Program for Soybean (NSF 9872565)

JOURNAL Unpublished (1999)

COMMENT On Oct 8, 1998 this sequence version replaced gi:3727950.
Other_ESTs: AI440994
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888)919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site:www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

```

FEATURES
source
    Location/Qualifiers
        1..796
            /organism="Glycine max"
            /cultivar="Williams"
            /db_xref="taxon:3847"
            /clone="Gm-r1021-1584"
            /clone_lib="Gm-r1021"
            /tissue_type="root"
            /lab_host="XL10-Gold"
            /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; Library Gm-r1021 is a sequence-driven, reracked set
of the original library Gm-cl004 which was prepared from
root cDNA. The mRNA was isolated from entire roots of 8
day old 'Williams' seedlings which were propagated on
paper towels with distilled water. Stratagene's cDNA
Synthesis Kit (catalog #200401) was used to synthesize
the cDNA. The Gm-cl004 library was constructed by Dr.
Paul Keim & Virginia H. Coryell, Department of Biology,
Box5640, Northern Arizona University, Flagstaff, AZ
86011, email: paul.keim@nau.edu, virginia.coryell@nau.edu.
The contig analysis to select unique genes was performed
by the laboratory of Ernest Retzel, Computational Biology
Centers, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and sequencing by the Keck
Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT      222 a    168 c    150 g    225 t    31 others
ORIGIN

```

BASE COUNT 222 a 168 c 150 g 225 t 31 others

Query Match 35.0%; Score 318.6; DB 71; Length 796;
Best Local Similarity 69.0%; Pred. No. 5.7e-67;
Matches 411; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Qy 6 acaatgagcactgcaagatttatcaagtgtgtcacggtcggtgatggagctgtggggaaa 65
 | | | | | | | | | | | | | | | | | | | | | | |
Db 794 AAAATGAGTGNNNNGANGTTTCATTAACTGCCTCNNNNTCNNCGACGTGCTGTCNNNAA 735

Qy 66 acttgtagtgcctattccatataccagcaatactttcccacgggattatgtttcaacagta 125
 | | } | | | | | | | | | | | | | | | | | | | |
Db 734 NNNNGCNNGTGAATTNNNACACCAGCAACACTTTCCACGGACTATGTGCCANNNTT 675

Qy 126 tttgaataactttagtccaatgtggtggtggtgagcgacagtgaaaccttggcctatgg 185

Email: dfrisch@CLEMSON.EDU

5 prime sequence.

FEATURES
source
Location/Qualifiers
1. 732
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande PtoR"
/db_xref="taxon:4081"
/clone="cLET19J17"
/clone_lib="tomato mixed elicitor, BTI"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XL1-Blue MRF"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLET - Inoculated with a variety of disease response
elicitors. Plants exposed to 2,6 dichloroisonicotinic
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
okadaic acid, or systemin prior to tissue harvest. EcoRI
site was destroyed during cloning."
BASE COUNT 213 a 119 c 194 g 206 t
ORIGIN

Query Match 37.1%; Score 337.6; DB 63; Length 732;
Best Local Similarity 74.1%; Pred. No. 1.5e-71;
Matches 427; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 5 aacaatgagcactgcaagatttatcaagtgtgtcacggtcggtgatggagctgtgggaa 64
Db 113 AAGAATGAGTGCTTCTAGGTTTATAAGTGTGTTACCGTGGGCGATGGAGCTGTGGGTAA 172
Qy 65 aactgtgatgtcatttcataaccagcaatactttcccaacggattatgttccaacagt 124
Db 173 AACTGTGCTCTCATTTCGTATACCAGCAACACTTTTCCCACTGATTATGTTCCCACTGT 232
Qy 125 atttgataacttttagtccaatgtgtgtgtgagtcagcagcagtgaaacctgtgctatg 184
Db 233 ATTTGACAATTTAGTGCAAAATGTGGTGTGCGATGGGAGCACTGTTAATCTGGGGCTCTG 292
Qy 185 ggacactgcccgggcaagaagattataataggctaaagccactgagttatagaggagctga 244
Db 293 GGATAGTCAGGTCAGGAGGATTACAATAGATTAGACCTTTGAGCTATCGTGGGGCTGA 352
Qy 245 tgtgtttttgttgcccttttctcttataagcaagccagttatgaaacatctacaaaaa 304
Db 353 TGTATTATAGTGGCATTTCCTCTCATTAGCAAGGCGAGCTATGAAATGTCTCCAAAAA 412
Qy 305 gtggatcccagagctaagacattatgtcataatgtaccagttgtgtgttggaaacaa 364
Db 413 GTGGATTCCGTAATTGAGGCATTATGCTCCTGGAGTTCCAATATTCTTGTGGAAACAA 472
Qy 365 actagatttgcgagatgacaagcagttcctcattgatcacctggagcaacaccaatc 424
Db 473 GCTAGATCTCCGAGAGGATAAGCAATCTTTTGTGGACATCCAGGTGCTGTTCCACTTAG 532
Qy 425 aacatctcaggagagaagaactaaagaagatgataggagcagttacttatatagaatgcag 484
Db 533 CACTGCTCAGGTCAGGAGCTGAGAAAGTCGATTGGTGTCTGCTTACATTGAATGTAG 592
Qy 485 ctccaaaacccaacagaatgtgaaggctgttttcgatgtgcaataaaagtagcttttag 544
Db 593 TGCAAAACCTCAACAGAACATTAGGCTGTTTTTGTGCGCCATTAGGTGGTCTCTACA 652
Qy 545 gccacaaaaccaaagagaagagccttgcaaaaggag 580
Db 653 ACCACCAAGCAAAAGAAGAAGAGAGGAGAAAGGG 688

RESULT 4
AW040005
LOCUS AW040005 732 bp mRNA EST 18-OCT-1999
DEFINITION EST282496 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
clone cLET19J17, mRNA sequence.
ACCESSION AW040005

VERSION AW040005.1 GI:5898759
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Asteridae; euasterids I; Solanales;
Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 732)
AUTHORS D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,
Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Niernan, W.,
Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and
Giovannoni, J.
TITLE Generation of ESTs from tomato leaf tissue
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188234.
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
3 prime sequence.

FEATURES
source
Location/Qualifiers
1. 732
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande PtoR"
/db_xref="taxon:4081"
/clone="cLET19J17"
/clone_lib="tomato mixed elicitor, BTI"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XL1-Blue MRF"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLET - Inoculated with a variety of disease response
elicitors. Plants exposed to 2,6 dichloroisonicotinic
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
okadaic acid, or systemin prior to tissue harvest. EcoRI
site was destroyed during cloning."
BASE COUNT 213 a 119 c 194 g 206 t
ORIGIN

Query Match 37.1%; Score 337.6; DB 63; Length 732;
Best Local Similarity 74.1%; Pred. No. 1.5e-71;
Matches 427; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 5 aacaatgagcactgcaagatttatcaagtgtgtcacggtcggtgatggagctgtgggaa 64
Db 113 AAGAATGAGTGCTTCTAGGTTTATAAGTGTGTTACCGTGGGCGATGGAGCTGTGGGTAA 172
Qy 65 aactgtgatgtcatttcataaccagcaatactttcccaacggattatgttccaacagt 124
Db 173 AACTGTGCTCTCATTTCGTATACCAGCAACACTTTTCCCACTGATTATGTTCCCACTGT 232
Qy 125 atttgataacttttagtccaatgtgtgtgtgagtcagcagcagtgaaacctgtgctatg 184
Db 233 ATTTGACAATTTAGTGCAAAATGTGGTGTGCGATGGGAGCACTGTTAATCTGGGGCTCTG 292
Qy 185 ggacactgcccgggcaagaagattataataggctaaagccactgagttatagaggagctga 244
Db 293 GGATAGTCAGGTCAGGAGGATTACAATAGATTAGACCTTTGAGCTATCGTGGGGCTGA 352
Qy 245 tgtgtttttgttgcccttttctcttataagcaagccagttatgaaacatctacaaaaa 304
Db 353 TGTATTATAGTGGCATTTCCTCTCATTAGCAAGGCGAGCTATGAAATGTCTCCAAAAA 412
Qy 305 gtggatcccagagctaagacattatgtcataatgtaccagttgtgtgttggaaacaa 364
Db 413 GTGGATTCCGTAATTGAGGCATTATGCTCCTGGAGTTCCAATATTCTTGTGGAAACAA 472
Qy 365 actagatttgcgagatgacaagcagttcctcattgatcacctggagcaacaccaatc 424

117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	379.2	41.7	662	79	AW690945	AW690945 NF034H11S
2	339.8	37.3	595	46	AI937960	AI937960 sc06b11.y
3	337.6	37.1	732	63	AW039993	AW039993 EST282484
4	337.6	37.1	732	63	AW040005	AW040005 EST282496
5	336.8	37.0	585	71	AW394676	AW394676 sh34a02.y
6	324.4	35.6	669	74	AW559248	AW559248 EST306084
c 7	318.6	35.0	796	71	AW349629	AW349629 GM210005B
8	308.4	33.9	649	44	AI759963	AI759963 sb66h11.y
9	307.2	33.8	688	45	AI900160	AI900160 sc01f12.y
10	306.8	33.7	592	80	AW705028	AW705028 sk41f03.y
11	306.6	33.7	555	45	AI901151	AI901151 sc21c12.y
12	305.6	33.6	658	45	AI900170	AI900170 sc01g12.y
13	304.2	33.4	549	45	AI901141	AI901141 sc21b12.y
14	303.4	33.3	713	44	AI759954	AI759954 sb66g11.y
15	299	32.9	625	79	AW688369	AW688369 NF006F02S
16	295.4	32.5	506	46	AI965741	AI965741 sc75d09.y
17	291	32.0	533	74	AW573665	AW573665 EST316256
18	290.2	31.9	549	36	AI162543	AI162543 A019P20U
19	282	31.0	606	46	AI941239	AI941239 sb86c10.y
20	278.6	30.6	485	74	AW573660	AW573660 EST316251
21	277.2	30.5	469	47	AU029919	AU029919 AU029919
22	267.2	29.4	517	80	AW705209	AW705209 sk43a11.y
23	266.4	29.3	622	43	AI727570	AI727570 BNLGH1842
24	264.8	29.1	410	74	AW559842	AW559842 EST314890
25	262	28.8	680	79	AW690086	AW690086 NF028B10S
26	261.2	28.7	435	44	AI812534	AI812534 12D8 Pine
27	257.2	28.3	437	69	AW202293	AW202293 sf13c10.y
28	255.4	28.1	638	43	AI731040	AI731040 BNLGH1845
29	252.2	27.7	463	48	AU082692	AU082692 AU082692
30	249.8	27.5	401	40	AI495724	AI495724 sb15e06.y
31	249.8	27.5	568	44	AI775563	AI775563 EST256663
32	247.2	27.2	696	80	AW694335	AW694335 NF075C06S
33	244	26.8	698	64	AW109094	AW109094 gate0002P
34	237.8	26.1	556	69	AW219991	AW219991 EST302474
35	236.6	26.0	551	79	AW621657	AW621657 EST312455
36	231.8	25.5	560	80	AW738459	AW738459 EST339886
37	229	25.2	688	79	AW685566	AW685566 NF031H02N
38	227	24.9	617	69	AW218480	AW218480 EST303663
39	225	24.7	524	43	AI730323	AI730323 BNLGH1662
c 40	224.6	24.7	621	62	AV440631	AV440631 AV440631
41	224.4	24.7	460	44	AI795130	AI795130 sb77d11.y
42	218.4	24.0	720	64	AW108575	AW108575 gate0001I
43	215.2	23.6	353	48	AU058227	AU058227 AU058227
44	214.6	23.6	378	40	AI460950	AI460950 sa78f02.y
c 45	212.8	23.4	616	47	AI999128	AI999128 701554548

ALIGNMENTS

RESULT 1
AW690945
LOCUS AW690945 662 bp mRNA EST 17-APR-2000
DEFINITION NF034H11ST1F1000 Developing stem Medicago truncatula cDNA clone
NF034H11ST 5', mRNA sequence.

ACCESSION AW690945
VERSION AW690945.1 GI:7565604
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 662)
AUTHORS Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research
JOURNAL Unpublished (2000)
COMMENT On Jan 6, 2000 this sequence version replaced gi:6676601.
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 662 Std Error: 0.00
Plate: 034 row: H column: 11
Seq primer: TCACACAGGAACAGCTATGAC.
FEATURES
source Location/Qualifiers
1..662
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF034H11ST"
/clone_lib="Developing stem"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/note="Vector: Lambda Zap; Contains a mixture of internodal stem segments"
BASE COUNT 207 a 100 c 140 g 211 t 4 others
ORIGIN

Query Match 41.7%; Score 379.2; DB 79; Length 662;
Best Local Similarity 81.9%; Pred. No. 1.3e-81;
Matches 435; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
Qy 2 aaaaacaatgagcactgcgaagatttatcaagtgtgtcacggtcggtgatggagctgtggg 61
Db 121 AAGAAGAATGAGTACTGCTAGGTTTATCAAGTGTGTAAACAGTAGGGGATGGTGTCTGG 180
Qy 62 gaaaactgtatgtctattcatataaccagcaatactttcccaacggattatgttccaac 121
Db 181 AAAAATTCGATGCTTATATCTTATACAGCAATACCTTTCCCACTGATTATGTTCCAAC 240
Qy 122 agtatttgataacttttagtgcgaatgtgtgtgtggatggcagcagcagtgaaacttggcct 181
Db 241 TGTGTTTGACAATTCAGTGCTAATGTAGTGGTGGATGTAGTACAGTTAATCTTGGTTT 300
Qy 182 atgggacactgccgggcaagaagtataataggtaagccactgagttatagaggagc 241
Db 301 ATGGGATACGCAGGACAAGAAGATTACAATAGATTAGGCCATTGAGTTACAGAGGAGC 360
Qy 242 tgatgtgtttttgtgtgctttctcttataagcaaggccagttatgaaacatctacaa 301
Db 361 TGATGTGTTTTTGTGTGTTTTCTCTCATAGTAAGCTAGTTATGAGAACATTCCAA 420
Qy 302 aaagtggatccagagcagaacattatgtctataatgtaccagttgtgtgttggaaac 361
Db 421 AAAGTGGATATCTGAGCTGAGCATTATGCTCCAAATGTGCTATTGTGCTGGTGGGAAC 480
Qy 362 caaactagatttgcgagatgacaagcagttctctattgatcaccctggaggaacaccaat 421
Db 481 AAAATTANATTTGCGAGCAGCAAGCAATTTTTATCGATCATCTGGAGCCACACAAT 540
Qy 422 atcaacatctcagggagaagaactaagaagatgataggagcagttactatataagaatg 481



Db 223 TGCAAGTGAAGGCAAACTGTGCACCTCCACATCTGGGACACAGGGCAAGAGTACT 282
Qy 208 ataataaggctaaggccactgagttatagaggagctgatgtgttttggccttttctc 267
Db 283 ATGACCGCTGCGGGCCCTGTCTACCTGACGCCAGCGTCTGCTGCTTTGCTTCGATG 342
Qy 268 ttataagcaaggccagtgtatgaaacatctacaaaagtggatccagagctaagacatt 327
Db 343 TCACCAGCCCGAACAGCTTTGACAACATCTTAACCGGTGGTACCCAGAGTGAATCATT 402
Qy 328 atgtcctaatagtaccagttgtgtgttgaaccaaactagattgagagatgacaa 385
Db 403 TCTGCAAGAAGGTACCATCATCGTCTGGGTGCAAGACTGACCTGCGCAAGGACAA 460

RESULT 13

US-08-766-551-2

; Sequence 2, Application US/08766551

; Patent No. 5840569

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Bandman, Olga

; APPLICANT: Hawkins, Phillip R.

; APPLICANT: Goli, Surya K.

; TITLE OF INVENTION: NOVEL HUMAN GTP-BINDING PROTEINS

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/766,551

; FILING DATE: Herewith

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0168 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 719 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: SEQ ID NO:2

; CLONE: 113700

US-08-766-551-2

Query Match 10.2%; Score 93; DB 3; Length 719;
Best Local Similarity 54.2%; Pred. No. 4e-16;
Matches 189; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 52 gagctgtggggaaaactgtatgctcatttcataaccagcaactttcccaaggatt 111
Db 33 GCGCGGTTTGGATGACAAACATTGGGGTGAGCTACACCCCAACGGCTACCCACCGAGT 92

Qy 112 atgttccaacagtatttgataactttagtccaatgtgtggtggatggcagcacagtga 171
Db 93 ACATCCCTACTGCCCTCGACAACCTTCTCCCGGTGGTGTCTGGATGGCGGCCCTGA 152
Qy 172 accttggcctatgggacactgcccgggcaagaagtataatagctaaggccactgagtt 231
Db 153 GACTCCAACCTCTGTGACACTGCCGACAGGATGAATTGACAAGCTGAGGCCCTCTCTGCT 212
Qy 232 atagaggagctgatgtgttttggccttttcttataagcaaggccagttatgaa 291
Db 213 ACACCAACACAGACATCTTCTGCTCTGCTTCAGTGTGTCGAGCCCTCATCTCCAGA 272
Qy 292 acatctacaaaaagtggatcccagactgaagcattatgctcataatgtaccagttgtgc 351
Db 273 ACGTCAGTGAGAAATGGGTGCCGAGATTCGATGCCACTGTCCAAAGCCCCCATCATCC 332
Qy 352 ttgttggaaaccaaactagatttgcgagatgacaagcagttcctcattga 400
Db 333 TAGTTGGAACGCAGTCGGATCTCAGAGAAGATGTCAAAGTCTCTATTGA 381

RESULT 14

US-08-247-946A-5

; Sequence 5, Application US/08247946A

; Patent No. 5792638

; GENERAL INFORMATION:

; APPLICANT: AARONSON, S.A.; CHAN, A.;

; APPLICANT: MIKI, T.

; TITLE OF INVENTION: NOVEL HUMAN RAS-RELATED

; TITLE OF INVENTION: ONCOGENES UNMASKED BY EXPRESSION OF cDNA

; TITLE OF INVENTION: CLONING

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: WORDPERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/247,946A

; FILING DATE: 24-MAY-1994

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: DOROTHY R. AUTH

; REGISTRATION NUMBER: 36,434

; REFERENCE/DOCKET NUMBER: 2026-4150

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 615

; TYPE: Nucleic acid

; STRANDEDNESS: Double

; TOPOLOGY: Unknown

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: No

; ORIGINAL SOURCE:

; ORGANISM: Human

; STRAIN:

; INDIVIDUAL ISOLATE:

; DEVELOPMENTAL STAGE:

; HAPLOTYPE:

; TISSUE TYPE:

; CELL TYPE:

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/842,976
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0267 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 702 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGN0T10
CLONE: 1379718
US-08-842-976-2

Query Match 11.5%; Score 105; DB 5; Length 702;
Best Local Similarity 55.7%; Pred. No. 2.5e-19;
Matches 201; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 28 tcaagtggttcacggctcggtgatgagactgtgggaaaaactgtatgctcatttcatata 87
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 TCAAGTGGTGCTCGGTGGGCGACGGCGGCTGCGGAAGACGTCGCTGCTGATGGTCTTCG 156

Qy 88 ccagacaatactttcccaacgattatgtccaacagtatttgataacttttagtgccaatg 147
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 CCGATGGGGCCTTCCCCGAGAGCTACACCCACGGTGTTTGAGCGGTACATGGTCAACC 216

Qy 148 tggtggtgatggtgcagcacagtaaaccttggectatgggacactgccggcggaagaagatt 207
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 217 TGCAAGTGAAAGGCAAACCTGTGCACCTCCACATGTGGGACACAGCAGGCGAGGAAGACT 276

Qy 208 ataataaggctaaggccaactgagttatatagaggagctgatgtgtttttgttgaccttttctc 267
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 ATGATCGACTGCGGCCCTCTCTCTACCCGGACACTGATGTCACTCATGTGCTTCTCCA 336

Qy 268 ttataagcaaggccagtttatgaaaacatctacaaaaagtggatcccagagctaagacatt 327
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 337 TCGACAGCCCTGACAGCCTGGAAAACATTCTTGAGAAGTGGACCCAGAGGTGAAGCACT 396

Qy 328 atgtctcataatgtaccagttgtgcttgttggaaccaaaactagatttgcgagatgacaagc 387
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 397 TCTGCCCAACGTGCCCATCATCTGTTGGGGAATAAGAAGACCTGAGGCAAGACGAGC 456

Qy 388 a 388
|
Db 457 A 457

RESULT 10
US-09-213-397-2
; Sequence 2, Application US/09213397
; Patent No. 6063377

```

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN RHO PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/213,397
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/842,976
; FILING DATE: 04/17/1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0267 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LUNGNOT10
; CLONE: 1379718
US-09-213-397-2

```

Query Match 11.5%; Score 105; DB 5; Length 702;
Best Local Similarity 55.7%; Pred. No. 2.5e-19;
Matches 201; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 128 tcaagtgtgtcacgctcggtgatggagctgtggggaaaactgtatgctcatttcataata 17
Db 97 TCAAGGTGGTCTGGTGGGCGACGGCGGCTGCGGGAAGACGTCGCTGCTGATGGTCTTCG 156

Qy 88 ccagcaataactttcccaacggattatgttccaacagattttgataacttttagtgccaatg 147
Db 157 CCGATGGGGCCTTCCCCGAGAGCTACACCCCCACGGTGTTTGAGCGGTACATGGTCAACC 216

Qy 148 tgggtggtgatgtggcagcacagtgaaccttggcctatgggacactgcccggcaagaagatt 207
Db 217 TGCAAGTGAAAGGCAAACTGTGCACCTCCACATGTGGGACACAGCAGGCAGGAAGACT 276

Qy 208 ataataaggctaaggccactgagttatagaggagctgatgtgtttttgttgcccttttctc 267
Db 277 ATGATCGACTGCGGCCTCTCTCTACCCGGACACTGATGTCTCCTCATCTGCTTCTCCA 336

Qy 268 ttataagcaaggccagttatgaaaacatctacaaaagtggatcccagagctaagacatt 327
Db 337 TCGACAGCCCTGACAGCCTGGAAAAACATTCTGAGAAGTGGACCCAGAGGTGAAGCACT 396

Qy 328 atgtgcataatgtaccagttgtgctttgtggaacaaactagattttgcgagatgacaagc 387
Db 397 TCTGCCCAACCTGCCCCATCTCTGGTGGGGAATAAGAAGGACCTGAGGCAGACGAGC 456

Qy 388 a 388

Qy 30 aagtgtgtcacggtcggtgatggagctgtgggaaaacttgatgtcatttcataacc 89
Db 19 aagctggtggtggtggcgacggcggtgtggcaagacgtgctgtgatcgtgttcagt 78
Qy 90 agcaatactttcccaacggattatgttccacagattttgataacttttagtccaatgtg 149
Db 79 aaggacgagttccccgaggtgtacgtgccacogtcttcgagaactatgtgcccagatt 138
Qy 150 gtgtgtggtgagcagcagtgaaaccttgccctatgggacactgcccgggcaagaagattat 209
Db 139 gaggtggacggcaagcaggtgagctggcgctgtgggacacggcgccagggagactac 198
Qy 210 aataggctaaggccaactgagttatagaggagctgatgtgtttttgttgccctttttctct 269
Db 199 gaccgctgctggcgctctctaccggacacogacgtcattctcatgtgctctcgtgtg 258
Qy 270 ataagcaaggccagttatgaaacatctacaaaagtggtccagagctaagaagattat 329
Db 259 gacagcccgactcgtggaacatccccgagaagtggtccccgaggtggaagcacttc 318
Qy 330 gctcataatgtaccagttgtgtgttggaacaaactagattgagagatgacaagcag 389
Db 319 tgtcccaatgtgcccacatcctgtgtgccaacaaaagacctgagcagcagcagcagc 378
Qy 390 ttcctcattgac 402
Db 379 gtcgcacagagc 391

RESULT 6

US-08-055-797-1

; Sequence 1, Application US/08055797

; Patent No. 5324830

; GENERAL INFORMATION:

; APPLICANT: RESNICK, MICHAEL A

; APPLICANT: CHOW, TERRY

; APPLICANT: PERKINS, ED

; TITLE OF INVENTION: A chimeric protein that has a human rho

; TITLE OF INVENTION: motif and deoxyribonuclease activity.

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN

; STREET: Eleventh Floor, 1615 L. Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20036-5601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/055,797

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/674,801

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: SCOTT, WATSON T

; REGISTRATION NUMBER: 26,581

; REFERENCE/DOCKET NUMBER: WTS/5683/83921/SRL

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3000

; TELEFAX: (202) 822-0944

; TELEX: 6714627CUSH

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2282 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 120..1574
US-08-055-797-1

Query Match 11.8%; Score 107.4; DB 1; Length 2282;
Best Local Similarity 58.3%; Pred. No. 8.9e-20;
Matches 208; Conservative 0; Mismatches 146; Indels 3; Gaps 1;

Qy 28 tcaagtgtgtcacggtcggtgatggagctgtgggaaaacttgatgtcatttcata 87
Db 334 TAAAGATTGTTGTTGGGAGATGGCGCTGTAGGGAAGACGTGCTGTATCTTATG 393
Qy 88 ccagcaatactttcccaacggattatgttccacagattttgataacttttagtccaatg 147
Db 394 TCCAAGGAACATTTCCGACTGATTATATCTTACTATTTTCGAAAATTATGTACAAACA 453
Qy 148 tgggtgtgtgagcagcagcagtgga--accttgccctatgggacactgcccgggcaagaag 204
Db 454 TAGAAGGACCCCAACGGTCAAATTTAGAAATGGCATTATGGGACACTGCCGGCCAGAAG 513
Qy 205 attataataggctaaggccactgagttatagaggagctgatgtgtttttgttgccctttt 264
Db 514 AGTATAGTAGACTTAGACCGCTTTTCATATAGGAATGCAGATGTCTGTGTTGCTATT 573
Qy 265 ctcttataagcaaggccagttatgaaacatctacaaaagtggtccagagctaagac 324
Db 574 CTGTTGGTAGTAGACATCGCTTAAAAATGTGAAGATCTCTGTTCCAGAGGTTAAGC 633
Qy 325 attatgctcataatgtaccagttgtgtgttggaacaaactagattgagagatg 381
Db 634 ATTTTGTCTTCCACTCCAATCATCTAGTCGGCCTTAAATCAGATCTATATGAAG 690

RESULT 7

US-07-914-284A-6

; Sequence 6, Application US/07914284A

; Patent No. 5489524

; GENERAL INFORMATION:

; APPLICANT: Chow, Terry Y.-K.

; APPLICANT: Resnick, Michael A.

; APPLICANT: Perkins, Edward

; TITLE OF INVENTION: A CHIMERIC PROTEIN THAT HAS A HUMAN RHO

; TITLE OF INVENTION: MOTIF AND DEOXYRIBONUCLEASE ACTIVITY

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson & Bear

; STREET: 620 Newport Center Drive, Sixteenth Floor

; CITY: Newport Beach

; STATE: CA

; COUNTRY: USA

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/914,284A

; FILING DATE: 14-JUL-1992

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/674,801

; FILING DATE: 26-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Altman, Daniel E.

; REGISTRATION NUMBER: 34,115

; REFERENCE/DOCKET NUMBER: NIH022.022CP1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (714) 760-0404

; TELEFAX: (714) 760-9502


```
Db 92 TCAAGGTGGTCTGGTGGGCGACGGCGGCTGCGGGAAGACGTGCTGCTGATGGTCTTCG 151
Qy 88 ccagcaataactttcccaacggattatgttccaacagtatttgataactttagtccaatg 147
Db 152 CGATGGGGCTTCCCCGAGAGCTACACCCACGGTGTGTTGAGCGGTACATGGTCAACC 211
Qy 148 tgggtggtggatggcagcagcagtgaaaccttggcctatgggacactgccgggcaagaagatt 207
Db 212 TGCAAGTGAAAGGCAAACCTGTGCACCTCCACATCTGGGACACAGCAGGGCAAGATGACT 271
Qy 208 ataataaggctaaggccactgagttatagaggagctgatgtgttttggccttttctc 267
Db 272 ATGACCGCCTGCGGCCCCCTGTCTACCTGACGCCAGCGTCTGCTGCTTTGCTTCGATG 331
Qy 268 ttataagcaaggccagttatgaaaacatctacaaaaagtggatccagagctaagacatt 327
Db 332 TCACCAAGCCCGAACAGCTTTGACAACATCTTTAACCGGTGGTACCCAGAAGTGAATCATT 391
Qy 328 atgtcataatgtaccagttgtgtgttggaaacaaactagatttgcgagatgacaa 385
Db 392 TCTGCAAGAAGGTACCCATCATCGTGGGCTGCAAGACTGACCTGCGCAAGGACAA 449
```

Search completed: September 3, 2000, 03:10:10
Job time: 28556 sec

```

RESULT 12
Q15017
ID Q15017 standard; DNA; 2282 BP.
AC Q15017;
DT 25-FEB-1992 (first entry)
DE Encodes yeast endo-exonuclease RhoNUC.
KW yeast cell cycle; rho/ras oncogene-like motif; RNC1 gene; ss.
OS Saccharomyces cerevisiae.

```

```

RESULT 13
V68232
ID V68232 standard; cDNA; 702 BP.
AC V68232;
DT 16-FEB-1999 (first entry)
DE Nucleotide sequence encoding human Rho.
KW ss; human; Rho protein; cell proliferation; inflammation;
KW transplantation; cancer; gene therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 45..663
FT /*tag= a
FT /product= "Human Rho"
PN W09846754-A1.
PD 22-OCT-1998.
PF 16-APR-1998; U07865.
PR 17-APR-1997; US-842976.
PA (INCY-) INCYTE PHARM INC.
PI Gol1 SK, Hillman JL;
DR WPI: 98-609916/51.

```

```
Qy      26 tatcaagtgtgctcacggctcggtagtgagactgtgtgggaacaaacctgatgtcatattcata 85  
        ||| | | | | | | | | | | | | | | | | |  
Db      279 TATAAAATGTGTTGTTGTCGGTGATGCGCTGGTAAAACCTTGCTTATTAACTCTCGTA 338  
  
Qy     86 taccagcaataacttcccacaggattatgtccaacagatatttgataacttttagtgcca 145  
        ||| | | | | | | | | | | | | | | | | |  
Db     339 TACCAGTAGTAAATTTCCAGCTGATTATGTCTACTGTTTTTGATAATTATGCTGTAAC 398  
  
Qy    146 tgttgtgtgtgtagtcgcagcagctgaaccttggcctatgggacactgccgggcaagaaga 205  
        ||| | | | | | | | | | | | | | | | | |  
Db    399 CGTGATGATAGGAGACAACCATTTACCTGGGGATTATTGATACTGCTGGTCAAGAAGA 458  
  
Qy    206 ttataatagctcaaggccactgagttagagaggctgatgtgtttttgtggccttttc 265  
        ||| | | | | | | | | | | | | | | | | |  
Db    459 TTACGACAGATTAAAGCCTTTTGTCATATCCATCGACTGATGTATCTCTGTTGTTTTTC 518  
  
Qy    266 tccttaagcacaagccagttatgaaaacatctacaaaaagtggatccccagctaagaca 325
```


Qy 430 ctcaggagagaagaactaaagaagatgataggagcagttacttatatagaatgcagctcca 489
Db 550 CTGAAGGCCAACAAAGTTGCTCAAGAATTGGTCTGCTGATTACTTGAATGTTCTGCTA 609
Qy 490 aaaccaacacagaatgtgaaggctgttttcgatgctgcaataaaagtagctttgagccac 549
Db 610 AAACCGGTAGAGGTGTAGAGAAAGTGTGAAGCTGCTACTAGAGCTTCTTAAGAGTTA 669
Qy 550' caaaaccaaagagaaag 566
Db 670 AAGAAAGAAAGGAAAG 686

RESULT 5

T92702

ID T92702 standard; cDNA; 3198 BP.
AC T92702;
DT 30-APR-1998 (first entry)
DE Candida Carhol gene.
KW GTPase; GGPTase; geranylgeranyl transferase; fungal Rho-like GTPase;
KW antifungal agent identification; mycosis; feed additive; disinfectant;
KW therapy; Candida cell detection; cell wall integrity; hyphal formation;
KW pathogenesis; Candida Carhol gene; ds.
OS Candida sp.
FH Key Location/Qualifiers
FT CDS 1362..1959
FT /*tag= a
PN WO9738129-A1.
PD 16-OCT-1997.
PF 10-APR-1997; U05929.
PR 10-APR-1996; US-631319.
PA (MITO-) MITOTIX INC.
PA (UIJO) UNIV JOHNS HOPKINS.
PI Berlin V, Damagnez V, Smith SE;
DR WPI; 97-512735/47.
DR P-PSDB; W30379.
PT Identification of antifungal agents that inhibit GTPase - useful for
PT specific detection of Candida
PS Claim 118; Fig 16; 123pp; English.
CC This sequence represents the Candida Carhol gene. The encoded protein is
CC a fungal Rho-like GTPase. The encoded protein can be used in an assay of
CC the invention. The method of the invention is for identifying potential
CC antifungal agents (I), and comprises: (a) mixing a fungal geranylgeranyl
CC transferase (GGPTase), a GGPTase substrate (A), and test compound;
CC and (b) detecting interaction between GGPTase and (A); a significant
CC reduction in this interaction indicates that the test compound is a (I).
CC (I) are useful for treating mycoses in humans or animals; as feed
CC additives and as disinfectants. This sequence, and Mab specifically
CC reactive with the encoded protein are used to detect Candida cells
CC specifically, particularly in cells, tissues and body fluids, while
CC antisense sequences are used to inhibit expression of these genes. The
CC method is a rapid, reliable and effective way of detecting agents that
CC inhibit GTPases, particularly those involved in cell wall integrity,
CC formation of hyphae and/or other cellular functions necessary for
CC pathogenesis. (I) should be selective for fungal cells, with little
CC effect on mammalian cells.
SQ Sequence 3198 BP; 1057 A; 538 C; 517 G; 1086 T;

Query Match 18.0%; Score 163.8; DB 1; Length 3198;
Best Local Similarity 57.3%; Pred. No. 6.8e-32;
Matches 319; Conservative 0; Mismatches 232; Indels 6; Gaps 1;

Qy 16 ctgcaagatttatcaagtggtcacggctggtgatggagctgtgggaaaactgtatgc 75
Db 1378 CTGAACCTTCGTAGAAAATTAGTCAATTGTCTGCTGCTTGTGGTAAGACTTGTAT 1437
Qy 76 tcatttcataaccagcaataactttcccaacggattatgttccacagattttgataact 135
Db 1438 TAAATGTTTTTCAAAGGTACTTCCAGAGTTTATGTCCCAACAGTTTTTGAATTT 1497
Qy 136 ttagtgccaatgtgtgtggtgatggcagcagctgaacettggcctatgggacactgccg 195

Db 1498 ACGTTGCTGATGTTGAAGTTGATGGTGAAGAAAGTTGATTTGGCATTATGGGATCTGCTG 1557
Qy 196 ggcaagaagattataataggtgaagccactgagttatagaggagctgatgtgtttttgt 255
Db 1558 GTCAAGAAGATTATGATAGATTAGACCATTATCTTCCAGATTCTAATGTTATTTTGA 1617
Qy 256 tggccttttctctataagcaaggccattatgaaacatctacaaaagtggatcccag 315
Db 1618 TTTGTTTTTCAGTTGATTCACAGATTCTTTAGATAACGTTTGAAGAAATGGATTCTG 1677
Qy 316 agctaagacattatgctcataatgtaccagtgtgtgtgttggaacaaactagatttgc 375
Db 1678 AAGTTTTTACATTTCTGTCAAGGTGTTCATCATTTTGTAGTTGGTTGAATCTGATTAA 1737
Qy 376 gagatgacaagcagttcct-----cattgatcacctggagcaaccaatatcaacat 429
Db 1738 GAGATGATCCTCATACTATTGAAGCCTTGAGACAACAACAACAACACCTCAACTT 1797
Qy 430 ctcaggagagaagaactaaagaagatgataggagcagttacttatatagaatgcagctcca 489
Db 1798 CTGAAGGCCAACAAAGTTGCTCAAGAATTGGTCTGCTGATTACTTGAATGTTCTGCTA 1857
Qy 490 aaaccaacacagaatgtgaaggctgttttcgatgctgcaataaaagtagctttgagccac 549
Db 1858 AAACCGGTAGAGGTGTAGAGAAAGTGTGAAGCTGCTACTAGAGCTTCTTAAGAGTTA 1917
Qy 550 caaaaccaaagagaaag 566
Db 1918 AAGAAAGAAAGGAAAG 1934

RESULT 6

T92869

ID T92869 standard; cDNA; 3198 BP.
AC T92869;
DT 30-APR-1998 (first entry)
DE Candida Carhol gene.
KW GTPase; GGPTase; geranylgeranyl transferase; fungal Rho-like GTPase;
KW antifungal agent identification; mycosis; feed additive; disinfectant;
KW therapy; Candida cell detection; cell wall integrity; hyphal formation;
KW pathogenesis; Candida Carhol gene; ds.
OS Candida sp.
FH Key Location/Qualifiers
FT CDS 1362..1959
FT /*tag= a
PN WO9738293-A2.
PD 16-OCT-1997.
PF 11-APR-1997; U05987.
PR 20-DEC-1996; US-771212.
PR 11-APR-1996; US-631319.
PA (MITO-) MITOTIX INC.
PA (UIJO) UNIV JOHNS HOPKINS.
DR WPI; 97-512864/47.
DR P-PSDB; W33896.
PT Identification of antifungal agents that inhibit GTPase - useful for
PT specific detection of Candida
PS Claim 118; Fig 16; 118pp; English.
CC This sequence represents the Candida Carhol gene. The encoded protein is
CC a fungal Rho-like GTPase. The encoded protein can be used in an assay of
CC the invention. The method of the invention is for identifying potential
CC antifungal agents (I), and comprises: (a) mixing a fungal geranylgeranyl
CC transferase (GGPTase), a GGPTase substrate GGPTase, and test compound;
CC and (b) detecting interaction between GGPTase and GGPTase; a significant
CC reduction in this interaction indicates that the test compound is a (I).
CC (I) are useful for treating mycoses in humans or animals; as feed
CC additives and as disinfectants. This sequence, and Mab specifically
CC reactive with the encoded protein are used to detect Candida cells
CC specifically, particularly in cells, tissues and body fluids, while
CC antisense sequences are used to inhibit expression of these genes. The
CC method is a rapid, reliable and effective way of detecting agents that
CC inhibit GTPases, particularly those involved in cell wall integrity,
CC formation of hyphae and/or other cellular functions necessary for

Db 121 CAGTATTGATAACTTTAGTGCCAATGTGGTGGATGGCAGCAGTGAACCTTGGCC 180
Qy 181 tatgggacactgcccggcaagaagattataatagctaaagccactgagttatagaggag 240
Db 181 TATGGGACACTGCCGGGCAAGAAGATTATATAGGCTAAGGCCACTGAGTTATAGAGGAG 240
Qy 241 ctgatgtgtttttgttgcccttttctcttaagaagccagttatgaaacatctaca 300
Db 241 CTGATGTGTTTTGTGGCCTTTTCTCTTATAGCAAGGCCAGTTATGAAACATCTACA 300
Qy 301 aaaagtggatcccagagctaaagacattatgctcataatgtaccagttgtctgttgaa 360
Db 301 AAAAGTGGATCCCAGAGCTAAGACATTATGCTCAATAATGTACCAGTTGTCTGTTGAA 360
Qy 361 ccaactagatttgcgagatgacaagcagttctcattgatccctggagcaacaccaa 420
Db 361 CCAACTAGATTGTGCGAGATGACAAGCAGTTCTCATTGATCACCTGGAGCAACACCAA 420
Qy 421 tatcaacatctcagggaagaactaaagaagatgtaggagcagttacttatagaat 480
Db 421 TATCAACATCTCAGGGAGAGAAGCTAAAGAAGATGATAGGAGCAGTTACTTATAGAAT 480
Qy 481 gcagctccaaaacccaacagaatgtgaagctgttttcgatgtgcaataaaagtagctt 540
Db 481 GCAGCTCCAAAACCCAACAGATGTGAAGCTGTTTTGCGATGTCGAATAAAGTAGCTT 540
Qy 541 tgaggcccaaaaaccaaagagaagccttgcaaaaggagaacatgtgttcccttgaa 600
Db 541 TGAGGCCCAAAAACCAAAGAGAAAGCCTTGCAAAAGGAGAACATGTGCTTTCCCTTGAA 600
Qy 601 tattggatcattattacagtcataaacagttacaaaagctgttgcagataaacactgaa 660
Db 601 TATTGGATCATTATTACAGTCAAAAACAGTTAACAAAAGCTGTGCGAGATAAACACTGAA 660
Qy 661 tctgctatagtttggttttgtttacatagttccacgtgaaactatgaagcatctctaa 720
Db 661 TCTGCTATAGTTTGGTTTGTGTTTACATAGTTCCACGTGAAACTATGAAGCATCTCTAA 720
Qy 721 gaaaacccaactatcatatcaaccatcgatcaatgaatgcatttcaatttgcagta 780
Db 721 GAAAACCCAACATCATATCAACCCATCGATCAATGAATGCATTCAATTTTGCAGTA 780
Qy 781 taagtctctttaaactccttttttacttcaatttataaagaattctatggataatgtt 840
Db 781 TAAGTCTCTTTAATCCTTTCTTTTACTTCATTTTATAACGAATCTATGGATAATGTT 840
Qy 841 cctacaaacatgtcattacaatgtttaattataaattccattctcttactataaaa 900
Db 841 CCTACAAACATGTCATTACAATGTTAATTATAAATCCATTCTCTATTACTAAAA 900
Qy 901 aaaaaaaaaa 910
Db 901 AAAAAAAAAA 910

RESULT 2

T73866

ID T73866 standard; DNA; 3045 BP.
AC T73866;
DT 26-JAN-1998 (first entry)
DE Cotton fibre promoter clone Rac13 construct, pCGN4735.
KW promoter; fibre-specific; transcriptional factor; promoter;
KW altered phenotype; colour; melanin; indigo; ss.
OS Gossypium hirsutum cv. coker 130.
PN WO9640924-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09897.
PR 07-JUN-1995; US-480178.
PR 01-JUL-1996; ZA-005572.
PA (CALJ) CALGENE INC.
PI McBride K, Pear JR, Perez-Grau L, Stalker DM;
DR WPI; 97-052325/05.

PT DNA construct contg. gene of interest controlled by cotton fibre
PT transcriptional factor - used to produce altered phenotype cotton
PT fibre cells expressing genes affecting pigmentation
PS Claim 23; Fig 5A-E; 95pp; English.
CC The present sequence is the Rac13 promoter construct, pCGN4735, isolated
CC from cotton fibre genomic clone 15-1. DNA constructs containing
CC cotton fibre-specific transcriptional factor promoters are useful to
CC produce cotton fibre cells with altered phenotype, especially altered
CC colour. Genes involved in the production of melanin (e.g. tyrosinase
CC gene and ORF438 encoded protein from Streptomyces antibioticus) and
CC indigo (mono-oxygenase genes possibly in conjunction with a
CC tryptophanase gene) are of interest. The promoters of the invention are
CC reliable and permit expression of a protein selectively in cotton fibre
CC to affect qualities such as fibre strength, length, colour and dyability
CC as required. The construct and methods can also be used for the
CC introduction of other advantageous genes into a cotton plant, e.g. a
CC plant hormone. In particular, fibres from a plant producing coloured
CC fibres may be used to produce yarns and/or fabrics that do not require
CC dyeing.
SQ Sequence 3045 BP; 1063 A; 450 C; 366 G; 1162 T;

Query Match 33.5%; Score 304.8; DB 1; Length 3045;
Best Local Similarity 97.8%; Pred. No. 4.6e-67;
Matches 309; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 594 ctttgatattggtatcattattacagtcataaacagtttaacaaaagctgttgcagataaa 653
Db 1806 CTCTGAATATTGGATCATTATTACAGTCAAAAACAGTTAACAAAAGCTGTTGCAGATAAA 1865
Qy 654 cactgaatctgctatagttgtttgtttgtttacatagttccacgtgaaactatgaagca 713
Db 1866 CACTGAATCTGCTATAGTTTGTGTTTGTGTTTACATAGTTCCACGTGAAACTATGAAGCA 1925
Qy 714 tctctaagaaaacccaactatcatatcaaccatcgatcaatgaatgcatttcaatttt 773
Db 1926 TCTCTAAGAAAACCCAACATCATATCAACCCATCGATCAATGAATCGATTTCATTTT 1985
Qy 774 cgcagtataagttcctttttaaactccttttcttttacttcaatttataaagaattctatgga 833
Db 1986 CGCAGTATAAGTTCCTTTTAACTCCTTTCTTTTACTTCATTTTATAACGAATTCATGGA 2045
Qy 834 taagtgtccctacaacatgtcattacaatgtttaattataaattccattctctatattt 893
Db 2046 TAAGTTCCTTCAAAACATGTCATTACAATGTTAATTATAAATCCATTCTCTATTTT 2105
Qy 894 actaaaaaaaaa 909
Db 2106 ACTAAGATATTAGTAA 2121

RESULT 3

T92698

ID T92698 standard; cDNA; 985 BP.
AC T92698;
DT 30-APR-1998 (first entry)
DE Candida CaRhol gene.
KW GTPase; GGPTase; geranylgeranyl transferase; fungal Rho-like GTPase;
KW antifungal agent identification; mycosis; feed additive; disinfectant;
KW therapy; Candida cell detection; cell wall integrity; hyphal formation;
KW pathogenesis; Candida CaRhol gene; ds.
OS Candida sp.
FH Key Location/Qualifiers
FT CDS 114..710
FT /*tag= a
PN WO9738129-A1.
PD 16-OCT-1997.
PF 10-APR-1997; U05929.
PR 10-APR-1996; US-631319.
PA (MITO-) MITOTIX INC.
PA (UYJO) UNIV JOHNS HOPKINS.
PI Berlin V, Damagnez V, Smith SE;
DR WPI; 97-512735/47.

```
Db 675 CAAACCAGGGAGAGAACTGAAGAACTGATTGGATCTGCTGTCTACATTGAATGTAGTT 734
Qy 487 ccaaaacccaacagaaatgtgaaggctgttttcgatgctgcaataaaagtagctttgaggc 546
Db 735 CAAAGACACAGCAGAACGTGAAGGCAGTGTGATGCAGCTATAAAAGTGGTGCTTCAGC 794
Qy 547 caccaaaaccaagagaaagccttgcaaaaggagaaacatgtgctttcctttga 599
Db 795 CACCAAAGCAGAAGAAGAAAAAATAAGAACCCTTGCGCGTTCTTGTA 847
```

Search completed: September 3, 2000, 04:03:47
Job time: 35777 sec

Db 826 ATCCCTGAGCTCAGACACTATGGCCATCGGTACCCATCTCTCGTTGGGACGAAGCTA 885
Qy 369 gatttgcgagatgacaagcagttcctcattgatcacctggagcaaccaatatcaaca 428
Db 886 GATCTTCGAGATGATAACAGTTCTTGTCTGACCATCTGGAGCGGCTCCAATTACAACC 945
Qy 429 tctcaggagagaactaagaagatgataggagcagttacttatatagaatgcagctcc 488
Db 946 TCTCAAGCGGAGGAGCTCAGGAAGTCAATTGGAGCGGCTCGTATATTGAGTGTAGCTCG 1005
Qy 489 aaaacccaacagaatgtgaaggctgttttcgatgctgcaataaaagtagctttgaggcca 548
Db 1006 AAGACTCAACAGAATGTGAAGCAGTTTTTGATGCAGCAATCAAGTGGTCTCTCAGCCA 1065
Qy 549 ccaaaaccaagagaaag 566
Db 1066 CCCAAGCAGAAGAAG 1083

RESULT 13

AF115476
LOCUS AF115476 1558 bp mRNA PLN 20-APR-1999
DEFINITION Physcomitrella patens rac-like GTP binding protein (rac2) mRNA,
complete cds.
ACCESSION AF115476
VERSION AF115476.1 GI:4588757
KEYWORDS
SOURCE Physcomitrella patens.
ORGANISM Physcomitrella patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;
Bryidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE 1 (bases 1 to 1558)
AUTHORS Winge,P., Kristensen,R., Bones,A.M. and Reski,R.
TITLE The Physcomitrella patens rac-gene family
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1558)
AUTHORS Kristensen,R., Winge,P., Bones,A.M. and Reski,R.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-1998) The Norwegian University of Science and
Technology, UNIGEN MTPS, Olav Kyrresgate 3, Trondheim N7005, Norway

FEATURES
source 1. .1558
/organism="Physcomitrella patens"
/db_xref="taxon:3218"
gene 1. .1558
/gene="rac2"
5'UTR 1. .530
/gene="rac2"
CDS 531. .1121
/gene="rac2"
/note="PhRac2"
/codon_start=1
/product="rac-like GTP binding protein"
/protein_id="AAD26198.1"
/db_xref="GI:4588758"
/translation="MSTSRFIKVTVDGAVGKTCMLISYTSNTFPTDYVPTVDNFS
ANVVVDGNTVNLGWDTAGQEDYNRLPLSYRGADVFLAFSLISKASYENISKWIP
ELRHYAPSVPIILVGTKLDRDQFFADHPGAAPITTSQSEELRKSIGAASYIECSS
KTQQNVKAVFDAIKVVLQPPKQKQKKKKQKNCVIL"
3'UTR 1122. .1558
/gene="rac2"
BASE COUNT 368 a 309 c 427 g 454 t
ORIGIN

Query Match 37.9%; Score 344.8; DB 8; Length 1558;
Best Local Similarity 72.7%; Pred. No. 5.6e-61;
Matches 445; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy 5 acaatgagcactgcaagatttatcaagtgtgtcacggctggtgatggagctgtgggaa 64
Db 527 AGCCATGAGCACTTCACGGTTTATCAAGTCGCTGACTGTTGGAGATGGAGCTGTCGGGAA 586

Qy 65 aactgtatgtcatttcataaccagcaactttcccaacggattatgttccaacagt 124
Db 587 GACGTGCATGCTTATTTATACACAGCAACACATTTCTACTGATTACGTCTCTACCGT 646
Qy 125 atttgataacttttagtccaatgtgtgtggtgagcagcagtagaaccttgccatag 184
Db 647 GTTTGACAACTCAGCGCAATGTAGTGGTCGATGGAATACCGTCAACCTCGGGTATG 706
Qy 185 ggacactgccgggcaagaagattataaggctaaggccactgagttatagaggagctga 244
Db 707 GGATACAGCAGGTCAAGAAGATTACAACAGGCTTCGCTCTGAGTTACAGGGGTGCTGA 766
Qy 245 tgtgtttttgttggccttttcttataagcaaggccagcttatgaaacatctacaaaa 304
Db 767 TGTTTTCTCTGGCGTTCTCCCTCATCAGCAAGGCTAGTTATGAAACATATCAAGAA 826
Qy 305 gtggatcccagagctaagacattatgctcataatgtaccagttgtgttgaaccaa 364
Db 827 GTGGATCCCGGAAGTACGACATTACGCGCCATCTGTCCAATCATTCTCGTGGAAACAA 886
Qy 365 actagatttgcgagatgacaagcagttcctcattgatcacctggagcaaccaatc 424
Db 887 ACTTGATCTTCGCGATGACAACAATTTCTTGTGATCCTCGGAGCGGCTCAATAAC 946
Qy 425 aacatctcagggagagaactaagaagatgataggagcagttacttatatagaatgcag 484
Db 947 TACTTCTCAAGGGGAGGAGCTCAGGAAGTCGATTGGGGCGGCTCGTACATAGAGTGCAG 1006
Qy 485 ctccaaaacccaacagaatgtgaaggctgttttcgatgctgcaataaaagtagctttgag 544
Db 1007 CTCAAAGACTCAGCAGAATGTAAAGCAGTTTTTGACGCAGCAATCAAGTGGTTCCTCA 1066
Qy 545 gccacaaaacccaagagaagagccttgcaaaaggagaacatgtgttcttccttgatatt 604
Db 1067 ACCACCAAGCAGAAGAAGAAGAAAAACAAAAGAAATTCGCTCATTCTGTGAATGTG 1126
Qy 605 ggatcattatta 616
Db 1127 GCATAGCTTTTA 1138

RESULT 14

NTA250174
LOCUS NTA250174 803 bp mRNA PLN 06-OCT-1999
DEFINITION Nicotiana tabacum mRNA for putative rac protein (rac gene).
ACCESSION AJ250174
VERSION AJ250174.1 GI:6015626
KEYWORDS rac gene; rac protein.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
Nicotiana.
REFERENCE 1 (bases 1 to 803)
AUTHORS Kieffer,F., Elmayan,T., Simon-Plas,F., Dagher,M.C. and Blein,J.P.
TITLE A tobacco cDNA encoding a Rac-like protein cloned using the
two-hybrid system in an heterologous screen
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 803)
AUTHORS Elmayan,T.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1999) Elmayan T., UMR 692 INRA/Universite de
Bourgogne, INRA / laboratoire de Phytopharmacie, BV 1540, 21034
Dijon Cedex, FRANCE

FEATURES
source 1. .803
/organism="Nicotiana tabacum"
/cultivar="Xanthi"
/db_xref="taxon:4097"
/dev_stage="55 day-old plants"
/tissue_type="young leaves: the nearest from apex"
/clone="Rac5"

```

RESULT 10
ATU64919
LOCUS ATU64919 1008 bp mRNA PLN 05-JAN-1999
DEFINITION Arabidopsis thaliana geranylgeranylated protein ATGP2 mRNA,
complete cds.
ACCESSION U64919
VERSION U64919.1 GI:4097562
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 1008)
AUTHORS Biermann,B.J., Price,J.R., Crowell,D.N. and Randall,S.K.
TITLE A collection of cDNAs encoding isoprenylated plant proteins
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1008)
AUTHORS Biermann,B.J., Price,J.R., Crowell,D.N. and Randall,S.K.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1996) Biology, IUPUI, 723 West Michigan Street,
Indianapolis, IN 46202-5132, USA
FEATURES
source Location/Qualifiers
1. .1008
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
CDS 223. .816
/note="similar to Rho1Ps; geranylgeranylated protein"
/codon_start=1
/product="ATGP2"
/protein_id="AAD00113.1"
/db_xref="GI:4097563"
/translation="MSASRPIKCVTVGDGAVGKTCLLISYTSNTFPTDYVPTVDFNF
ANVYVNGATVNLGLWDTAGQEDYNRLRPLSYRGADVFIILAFSLISKASYENVSKKWI
ELKHYPAGPVIPLVGTGKILDRDDQKFIFDHGPAVPIITTAQGEEKKLIGAPAYIECS
RTQENVGVFDAAIRVVLQPPKQKKKSKAQKACSIL"
BASE COUNT 276 a 191 c 221 g 320 t
ORIGIN

```

Qy 2 aaaaacaatgagcactgcaagatttatcaagtgtgtcacggctcggtgatggagctgtggg 61
Db 216 AAGAGAAATGAGCGCTTCGAGGTTCTATAAGTGTGTCAACGTTGGCGACGGAGCTGTGG 275

Qy 62 gaaaacttgtagtctcatttcataataccagcaatactttcccaacggattatgttccaac 121
Db 276 TAAAACTGTTTGCTGATTCTTACACAGCAACACTTTTCCTACGGATTATGTACCGAC 335

Qy 122 agtatttgataacttttagtgccaatgtggtggtgatggcgcagcacagtgaaacctgtgacct 181
Db 336 TGTTTTCGATAACTTTAGCGCAAATGTGGTGTGTAATGGAGCCACTGTGAATCTGGGCCT 395

Qy 182 atgggacactgcccggcgaagaagattataataggctaaggccactgagttatagaggagc 241
Db 396 ATGGGATACCGCAGGCGCAGGAGGATTATAACAGATTAAGACCTTTGAGTTACCGCGGTGC 455

Qy 242 tgatgtgttttgtgtggcctttttctcttataagcaaggccagttatgaaaacatctacaa 301
Db 456 TGATGTTTTTCATCTTAGCATTCTCTCTATCAGTAAGGCTAGTTATGAGAATGCTCCAA 515

Qy 302 aaagtggatcccagagctaagacattatgtctcataatgtaccagttgtgcttgttggaa 361
Db 516 GAAGTGGATCCCAGAGCTGAAGCATTATGCCCTGGTGTCCCTATAGTCTTGTGTGGAAC 575

Qy 362 caaactagattttgcgagatgacaagcagttctctcattgatcaccttggagcaacaccaat 421
Db 576 CAAACTAGATCTTCGGGATGACAAACAGTCTTCTATTGACCACCTTGGCGCTGTACCAAT 635

Qy 422 atcaacatctcaggggagaagaactaaagaagatgataggagcagttacttatatagaatg 481
Db 636 TACTACTGCTCAGGGAGAGGAACTGAAGAACTAATTGGAGCTCCGCATACATCGAGTG 695

Qy 482 cagctccaaaacccaacagaatgtgaaggctgttttctgatgtgcaataaaagtagcttt 541
Db 696 CAGTTCAAAACACAGAGAACGTAAGAGGAGTATTGTATGCAGCGATCGAGTGGTTCT 755

Qy 542 gagggccacaaaaccaaagagaagaagccttgcaaaaggagaacatgtgctttctcttgaat 601
Db 756 TCAACCTCCAAGCAGAGAAAAAGAAAGCAAGCAAAAAAGCCTGCTCCATTTTGTA 815

Qy 602 att 604
Db 815 ATT 818

```

RESULT 11
ATU41295
LOCUS ATU41295 1191 bp mRNA PLN 28-OCT-1997
DEFINITION Arabidopsis thaliana GTP binding protein (ARAC1) mRNA, complete
          cds.
ACCESSION U41295
VERSION U41295.1 GI:1292907
KEYWORDS
SOURCE
  thale cress.
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE 1 (bases 1 to 1191)
AUTHORS Winge,P., Brembu,T. and Bones,A.M.
TITLE Cloning and characterization of rac-like cDNAs from Arabidopsis
       thaliana
JOURNAL Plant Mol. Biol. 35 (4), 483-495 (1997)
MEDLINE 98009984
REFERENCE 2 (bases 1 to 1191)
AUTHORS Winge,P., Brembu,T. and Bones,A.

```

RESULT 7
ATU49972
LOCUS ATU49972 . 935 bp DNA PLN 19-NOV-1998
DEFINITION Arabidopsis thaliana GTP binding protein Rop2At (Rop2At) mRNA,
complete cds.
ACCESSION U49972
VERSION U49972.1 GI:1777763
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 935)
AUTHORS Li,H., Wu,G., Ware,D., Davis,K.R. and Yang,Z.
TITLE Arabidopsis rho-related GTPases: differential gene expression in
pollen and polar localization in fission yeast
JOURNAL Plant Physiol. 118 (2), 407-417 (1998)
MEDLINE 98440662
REFERENCE 2 (bases 1 to 935)
AUTHORS Li,H., Lin,Y., Ware,D., Zhou,D., Davis,K.R., Cramer,C.L. and
Yang,Z.
TITLE Differential Developmental Expression of Two Arabidopsis Genes
Encoding Rho Family Small GTPases
JOURNAL Unpublished

[illegible]

Db 81 AAAGACTTGTATGCTCATTTCATATACCAGCAATACGTTTCTACGGATTATGTTCCAAC 140
Qy 122 agtatttgataacttttagtgccaatgtggtggtgagtcagcagcagtgaaaccttgacct 181
Db 141 AGTTTCGACAACTTCAGCGCAATGTGGTGGTGCAGCGGAGTACCGTGAACCTTGGCCT 200
Qy 182 atgggacactgccgggcaagaagattataataggctaaggccactgagttatagaggagc 241
Db 201 GTGGGATCTGCCGGTCAGGAAGATTATAATAGGCTTAGGCCCTTGAGTTACAGAGGAGC 260
Qy 242 tgatgtgtttttgttgaccttttcttataagcaaggccagttatgaaacatctacaa 301
Db 261 AGATGCTCTTATTAGCATTTTCCCTTATAAGCAAGGCCAGTTACGAGAATATTACAA 320
Qy 302 aaatggatccagagctaagacattatgctcataatgtaccagttgtgcttgggaac 361
Db 321 AAATGGGCTTCCGGAGCTGAACATTATGCTCTGGCATCCCATTTGCTCTCGGAAC 380
Qy 362 caaactagatttgcagatgacaagcagttcctcattgatcacctggagcaacaccaat 421
Db 381 AAAATTAGATTGAGGGATGACAGCAGTCTCTGAAGGATCATCCAGGACAGTCTCTAT 440
Qy 422 atcaacatctcagggagaagaactaaagaagatgataggagcagttacttatatagaatg 481
Db 441 AACAACTGCTCAGGGAGAAGAAATTAAGAAAATGATTGGAGCTGTAGGTACTTAGAGTG 500
Qy 482 cagctccaaaacccaacagaatgtgaaggctgttttcgatgctgcaataaaagtagcttt 541
Db 501 CAGCTCCAAAACCAACAGAATGTGAAGGCAGTGTGTGATACAGCGATAAGGGTAGCTTT 560
Qy 542 gagggccacaaaacaaagagagaag 566
Db 561 GAGGCCACCAAGGCAAGAAAAAG 585

RESULT 4
LJRAC2
LOCUS LJRAC2 982 bp mRNA PLN 12-MAY-1997
DEFINITION L.japonicus mRNA for small GTP-binding protein, RAC2.
ACCESSION 273962
VERSION 273962.1 GI:1370200
KEYWORDS rac2 gene; small GTP-binding protein.
SOURCE Lotus japonicus.
ORGANISM Lotus japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE 1 (bases 1 to 982)
AUTHORS Borg, S., Brandstrup, B. and Poulsen, C.
TITLE Structural analysis of cDNAs encoding 33 different small GTP
binding proteins from Lotus japonicus and expression of
corresponding mRNAs in developing root nodules
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 982)
AUTHORS Poulsen, C.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1996) C. Poulsen, University Of Aarhus, Dept of
Molecular and Structural Biology, Gustav Wieds Vej 10C, DK-8000
Aarhus C, DENMARK
REFERENCE 3 (bases 1 to 982)
AUTHORS Borg, S., Brandstrup, B., Jensen, T.J. and Poulsen, C.
TITLE Identification of new protein species among 33 different small
GTP-binding proteins encoded by cDNAs from Lotus japonicus, and
expression of corresponding mRNAs in developing root nodules
JOURNAL Plant J. 11 (2), 237-250 (1997)
MEDLINE 97231679
FEATURES
source Location/Qualifiers
1. .982
/organism="Lotus japonicus"
/variety="Gifu B-129"
/db_xref="taxon:34305"
/tissue_type="root nodules"

/dev_stage="21 dpi with Rhizobium loti NZP 2037"
/clone_lib="lambda ZAPII (Stratagene)"
gene 113. .703
/gene="rac2"
CDS 113. .703
/gene="rac2"
/function="GTP-binding protein"
/codon_start=1
/product="RAC2"
/protein_id="CAA98190.1"
/db_xref="GI:1370201"
/db_xref="SWISS-PROT:Q40220"
/translation="MSTARPIKCVTVGDGAVGKTCMLISYTSNTFPTDYPVTFDNFS
ANVVVGSTVNLGLWDTAGQEDYNLRPLSYRGADVFLAFSLLSRASYENISKWIP
ELRHYAPTVPVILVGTGLDLREDQYLDHPGATPITTAQGEELKAIGAAYVLECS
KTQNVKAVFDAIKVVLQPPKPKKKRKRTRPCVFL"
BASE COUNT 290 a 173 c 211 g 308 t
ORIGIN

Query Match 43.6%; Score 396.8; DB 7; Length 982;
Best Local Similarity 78.8%; Pred. No. 1.4e-71;
Matches 473; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 6 acaatgagcactgcaagatttatcaagtgtgtcagcggtcggtgaggtgtggtggaa 65
Db 110 AAAATGAGCAGCAGTAGATTCATCAAGTGTGTACTGTGGAGATGGAGCAGTGGGAAG 169
Qy 66 actgtatgctcatttcataaccagcaatacttcccaacggattatgtccaacagta 125
Db 170 ACCGTATGCTTATCTTACACCAGCAACACATTCACCAGGATTATGTGCTACTGTT 229
Qy 126 ttgtataacttttagtgccaatgtggtggtgagtcagcagcagtgaaaccttgacctatgg 185
Db 230 TTTGATACTTCAGTGCAGTGTGGTGTGATGGCAGCAGTAACTGGGATTATGG 289
Qy 186 gacactgccgggcaagaagattataataggctaaggccactgagttatagaggagctgat 245
Db 290 GACACTGCTGGACAGGAGGATTACAATAGGCTTAGGCCCTTGAGCTACAGAGGAGCAGAT 349
Qy 246 gtgtttttgttgaccttttcttataagcaaggccagttatgaaacatctacaaaag 305
Db 350 GTGTCTTCTGCTGGCTTTTCCCTCCTTAGCAGGCCAGCTATGAAATATCTCCAAAAG 409
Qy 306 tggatccagagctaagacattatgctcataatgtaccagttgtgcttgggaaccaa 365
Db 410 TGGATTCTGAAGTACAGACTATGCCCCAATGTCGAATGTCTTGTGGGAACCAA 469
Qy 366 ctgatttgcagatgacaagcagttcctcattgatcacctggagcaacaccaatata 425
Db 470 CTTGATTGAGGGAAGCAGGAGCAGTATTGATTGATCATCTGGAGCCACCATTTACT 529
Qy 426 acatctcagggagaagaactaaagaagatgataggagcagttacttatatagaatgcagc 485
Db 530 ACTGCCCGGAGAAGAGCTGAAGAAGCAATGGTGTGCTGTGTACCTAGAAATGCAGC 589
Qy 486 tccaaaacccaacagaatgtgaaggctgttttcgatgctgcaataaaagtagctttgagg 545
Db 590 TCAAAGACTCAACAGAAATGTGAAGGCTGTGTTGATGCTGCTATCAAGGTTGTTTGCAG 649
Qy 546 ccacaaaacccaagagaagccttgcaaaaggagaacatgtgcttctcttgaatattg 605
Db 650 CCACCTAAACCAAGAAAAACGAAGAAGACCAACCATGCGTTTTCTTTAATTATGATG 709

RESULT 5
ATU49971
LOCUS ATU49971 843 bp DNA PLN 19-NOV-1998
DEFINITION Arabidopsis thaliana GTP binding protein Rop1at (Rop1at) mRNA,
complete cds.
ACCESSION U49971
VERSION U49971.1 GI:2558665
KEYWORDS

25	334.4	36.7	1008	8	NTU64924	U64924 Nicotiana t
26	333.4	36.6	1117	8	AF031428	AF031428 Arabidops
27	333	36.6	1220	8	AF051223	AF051223 Picea mar
28	332.6	36.5	917	8	MSA251210	AJ251210 Medicago
29	330.8	36.4	850	49	AF218381	AF218381 Oryza sat
30	329.8	36.2	1059	7	AB024996	AB024996 Cicer ari
31	326	35.8	756	8	AFU64920	U64920 Arabidopsis
32	325.8	35.8	1081	49	AF233447	AF233447 Physcomit
33	320.4	35.2	1393	8	AF126053	AF126053 Zea mays
34	317.2	34.9	1058	8	AF126055	AF126055 Zea mays
35	313.8	34.5	869	8	AF079486	AF079486 Arabidops
36	310.6	34.1	640	49	AF165925	AF165925 Gossypium
37	308	33.8	771	8	AF156896	AF156896 Arabidops
38	291.8	32.1	1067	7	AB029510	AB029510 Oryza sat
39	290.8	32.0	1045	8	AF126054	AF126054 Zea mays
40	290	31.9	794	8	AF079485	AF079485 Arabidops
41	286	31.4	996	7	AB029508	AB029508 Oryza sat
42	285.8	31.4	867	49	AF239751	AF239751 Tradescan
43	283.2	31.1	1090	8	AF079484	AF079484 Arabidops
44	282.2	31.0	1127	8	AF126052	AF126052 Zea mays
45	277.6	30.5	1087	7	AB029509	AB029509 Oryza sat

ALIGNMENTS

RESULT 1
S79308
LOCUS S79308 913 bp mRNA PLN 30-NOV-1995
DEFINITION Rac13=21.8 kda GTP-binding protein [Gossypium hirsutum=cotton plants, cv. Acala SJ-2, boll fibers, mRNA Partial, 913 nt].
ACCESSION S79308
VERSION S79308.1 GI:1087110
KEYWORDS
SOURCE upland cotton boll fibers cv. Acala SJ-2.
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 913)
AUTHORS Delmer,D.P., Pear,J.R., Andrawis,A. and Stalker,D.M.
TITLE Genes encoding small GTP-binding proteins analogous to mammalian rac are preferentially expressed in developing cotton fibers
JOURNAL Mol. Gen. Genet. 248 (1), 43-51 (1995)
MEDLINE 95379748
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 170155] from the original journal article.
This sequence comes from Fig. 1A.

FEATURES
source Location/Qualifiers
1..913
/organism="Gossypium hirsutum"
/db_xref="taxon:3635"
gene 12..602
/gene="Rac13"
/note="21.8 kda GTP-binding protein"
CDS 12..602
/gene="Rac13"
/note="21.8 kda GTP-binding protein; pea Rho1 protein homolog/mammalian rac protein homolog; ; This sequence comes from Fig. 1A"
/codon_start=1
/protein_id="AAB35093.1"
/db_xref="GI:1087111"
/translation="MSTARFIKCVTVGDVAVGKTCMLISYTSNFTDYPVTFVDFNS ANVVVDGTVNLGLWDTAGQEDYNRLRPLSYRGADVFLAFSLISKASYENIYKKWIP ELRHYAHNVFVVLVGTGLDLRDDKQPLIDHPGATPISTSQGEELKMGIAVITYIECSS KTOQNKAVFDAAIKVALRPPKPKRRPKRRTCAPL"

BASE COUNT 307 a 169 c 172 g 265 t
ORIGIN

Query Match 100.0%; Score 910; DB 8; Length 913;
Best Local Similarity 100.0%; Pred. No. 1.7e-176;

Matches 910; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	aaaaaacatgagcactgcaagatttatcaagtgtgtcaggtcggtgatggagctgtgg	60
Db	4	AAAAACAATGAGCACTGCAAGATTATCAAGTGTGTACGGTCGGTGATGGAGCTGGG	63
Qy	61	ggaaaactgtatgctcatttcataaccagcaatacttcccaacggattatgttcca	120
Db	64	GGAAAACCTGTATGCTCATTTCATATACCAGCAATACTTTCCCAACGGATTATGTCCAA	123
Qy	121	cagtattgataacttttagtccaatgtgtgtgtggatggcagcacagtgaaccttgccc	180
Db	124	CAGTATTGTATACTTTAGTGCCAATGTGGTGGTGGATGGCAGCAGTGAACCTTGGCC	183
Qy	181	tatgggacactgcccgggaagattataataggctaaggcactgagttatagaggag	240
Db	184	TATGGGACACTGCCGGCAAGAAGATTATAATAGGCTAAGGCCACTGAGTTATAGAGGAG	243
Qy	241	ctgatgtgtttttgttgcccttttctcttataagcaaggcagttatgaaacatctaca	300
Db	244	CTGATGTGTTTTGTGTGGCCTTTTCTCTTATAAGCAAGGCCAGTTTAAACATCTACA	303
Qy	301	aaaagtggatccagagctaagacattatgctcataatgtaccagttgtgcttgttggaa	360
Db	304	AAAAGTGGATCCAGAGCTAAGACATTATGCTCATAATGTACCAGTTGTCTTGTGGAA	363
Qy	361	ccaaactagatttgcgagatgacaagcagttctctcattgatccacctggagcaacaccaa	420
Db	364	CAAACTAGATTTCGAGAGTGAACAGCAGTTCCTCATTGATCACCTGGAGCAACACCAA	423
Qy	421	tatcaacatctcagggaagaagaactaaagaagatgataggagcagttacttatatagaat	480
Db	424	TATCAACATCTCAGGGAGAAGAACTAAAGAAGATGATAGAGCAGTTACTTATATAGAAT	483
Qy	481	gcagctccaaaacccaacagaatgtgaaggctgttttcgatgtgcaataaaagtagctt	540
Db	484	GCAGCTCCAAAACCAACAGAATGTGAAGGCTGTTTCGATGTGCAATAAAAGTAGCTT	543
Qy	541	tgaggccacaaaaccaaagagaagccttgcaaaaggagaacatgtgctttccttggaa	600
Db	544	TGAGGCCACCAAAACCAAGAGAAAGCCTTGCAAAAGGAGAACATGTGCTTTCCITTGAA	603
Qy	601	tattggatcattattacagtcacaaacagtttaacaaagctgttgagataaacactgaa	660
Db	604	TATTGGATCATTATTACAGTCAAAACAGTTTAACAAAGCTGTTCAGATAAACACTGAA	663
Qy	661	tctgctatagttgttttggtttacatattgtccacgtgaaactatgaagcatctctaa	720
Db	664	TCTGCTATAGTTGTTTTTGGTTACATATGTCCACGTGAACATATGAGCATCTCTAA	723
Qy	721	gaaaacccaactatcatatcaaccatcgatcaatgaatcatttcaatttcgcagta	780
Db	724	GAAAACCAACTATCATATCAACCATCGATCAATGAATCGATTTCAAATTTTCGAGTA	783
Qy	781	taagtctcttttaactcttttcttttacttcttttataacgaattctatggataatgtt	840
Db	784	TAAGTCTCTTTAATCTTTCTTTTACTTCTATTATAACGAATTCATGATAATGTT	843
Qy	841	ccctacaacatgtcattacaatgtttaattataaattccattctcttatttactaaaa	900
Db	844	CCCTACAACATGTCATTACAATGTTTAATTATAAATTCATTCTCTATTTTACTAAAA	903
Qy	901	aaaaaaaaa 910	
Db	904	AAAAAAAAA 913	

RESULT 2
S79309
LOCUS S79309 840 bp mRNA PLN 30-NOV-1995
DEFINITION Rac9=21.5 kda GTP-binding protein [Gossypium hirsutum=cotton plants, cv. Acala SJ-2, boll fibers, mRNA Partial, 840 nt].
ACCESSION S79309

1

Qy 169 aaatcagaaagcacgaagagtctgatacaaacaggccaaaatatcatgaagagtacc 228
||| | ||| | | | | | | | | | |
Db 1092 AAAAAAAAAAANNNNNAAAAAAAAAANNAANNAANNAANNA 1033

Qy 229 aaacatgagaagcctgaagtgtacacaggaggaacaaacacctgcacaatcatgaa 288
||| | | | | | | | | | | | | | | | |
Db 1032 AAAAAAAAAAANNAANNAAAAAAAAAANNAANNAANNAANNA 973

Qy 289 gagtaccacgagtcacycgaaatcgaaggagcacgaagagtacgataaagaaaaaccgat 348
| | | | | | | | | | | | | | | | | |
Db 972 AAAAAAAAAANNAANNAANNAAAAAAAAAAAAAAAAAAAAAAA 913

Qy 349 ttccccaatgggaaagcctaagaagcacgagaacacgaagtgcgaatatccgaaata 408
| | | | | | | | | | | | | | | | | |
Db 912 NNAAAAAAAAANAAAAAAAAAAAAAAAAANAAAAAAAAANAAAA 853

Qy 409 cccgagtacaaggacaacaaagatgagaataagaacaataagatgaagagtgccaggag 468
| | | | | | | | | | | | | | | | | |
Db 852 AAANAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 793

Qy 469 tcacacgaatcgaagagcacgaagagtacgagaagaaaaaccgatttccccaatgg 528
| | | | | | | | | | | | | | | | | |
Db 792 AAANAAAAAAAAAAAAAAAAAAAAAAAAANNAANNAANNAANNA 733

Qy 529 gaaaagcctaagggccacgagaacataaagccgaatatccgaaaatacctgagtgcag 588
||| | | | | | | | | | | | | | | |
Db 732 AAAAAAAAAAANNNNNAAAAAAAAAAAAAAAAAAAAAAAAAANNA 673

Query Match 8.9%; Score 85.8; DB 101; Length 870;
Best Local Similarity 44.6%; Pred. No. 8.1e-11;
Matches 225; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

```

RESULT      7
AQ782441
LOCUS      AQ782441      693 bp      DNA      GSS      02-AUG-1999
DEFINITION HS_3174_A2_B03_MR CIT Approved Human Genomic Sperm Library D
Homo sapiens genomic clone Plate=3174 Col=6 Row=C, genomic
survey sequence.
ACCESSION  AQ782441
VERSION    AQ782441.1 GI:5685401
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 693)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    On Mar 23, 1999 this sequence version replaced gi:3324197.
            Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3174 row: C column: 6
            Seq primer: M13 Reverse

```

Qy 168 aaatacgaataaacgacgaagagtctgtaatacaaacagcccaaatatcatgaagagtacc 227
 Db 175 AAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAANNN 234

Qy 228 aaacatgagaagcctgaatgtacagaggagaaaaacaaacccctgcaacatcatga 287
 Db 235 AAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAANNN 294

Qy 288 agagtaccacgagtcgcggaatcgaaggagcagcaagagtacgataagaaaaacccga 347
 Db 295 AAAAAAAAAANAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAA 354

Qy 348 ttcccccaaatgggaagcctaagagcagcagaaacacgaagtctgaatatccgaatat 407
 Db 355 AANNANNAANNAANNNNAAAAAAAAANAAAAAAAAAAAAAAAAAAAA 414

Qy 408 acccgagtacaggacaacaagaatgagataagaacataaagatgaagagtgccagga 467
 Db 415 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAANNAAN 474

Qy 468 gtcacacgaatcgaaggagcagcaagagtagtcagagaagaaaaacccgattcccccaatg 527
 Db 475 NAAAAAAAAAAAAAAAAANNAANAAAAAAAAAAAAAAAAANAAAAAAAA 534

Qy 528 ggaagaagcctaagaaggcagcagagaacataaagccgaatatccgaaatctctgagtcaa 587
 Db 535 AAAAAAAAAAAAAAAAAANANAAAAAAAAAAAAAAAAAAAAAAAAAN 594

Qy 588 ggaaaactagatgaggataaggaacataaacatgagtcccaagcatgaaagaaga 647
 Db 595 ANANANAAAAAAAAAAAAAAAAAAAAAAAAANNAAAAAAAAAAAAAAA 654

Qy 648 ggagaagaaa 657
 Db 655 AAAAAAAAAAA 664

```

RESULT      8
CNS0122R
LOCUS       CNS0122R      839 bp      DNA              GSS              26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
              BAC07E20 of DrosBAC library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION   ALL01037
VERSION     AL101037.1  GI:5612648
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 839)
AUTHORS     Genoscope.
TITLE       Direct Submission

```

RESULT 4
B12963/c
LOCUS B12963 759 bp DNA GSS 14-MAY-1997
DEFINITION T23D1-T7.1 TAMU Arabidopsis thaliana genomic clone T23D1,
genomic survey sequence.
ACCESSION B12963
VERSION B12963.1 GI:2094085
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 759)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT On Dec 15, 1999 this sequence version replaced gi:4123328.
Other_GSSs: T23D1-Sp6.1, T23D1-Sp6, T23D1-T7
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 88
High quality sequence stop: 127.
FEATURES Location/Qualifiers

```

RESULT      5
CNS00HGZ
LOCUS       CNS00HGZ      1101 bp      DNA              GSS              03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
             BACR35012 of RPCI-98 library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION   AL073472
VERSION     AL073472.1   GI:4953252
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster

```

117: gb_gss13:*
118: gb_gss14:*
119: gb_gss15:*
120: gb_gss16:*
121: gb_gss17:*
122: gb_gss18:*
123: gb_gss19:*
124: em_gss13:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
	1	95.8	9.9	593	122	CNS00880	AL051540 Drosophil
c	2	94.2	9.7	997	122	CNS005TE	AL060767 Drosophil
c	3	93.8	9.7	796	122	CNS0118D	AL099943 Drosophil
c	4	90.8	9.4	759	120	B12963	B12963 T23D1-T7.1
	5	90.6	9.4	1101	122	CNS00HGZ	AL073472 Drosophil
	6	88.2	9.1	952	123	CNS014BF	AL103941 Drosophil
	7	87.8	9.1	693	114	AQ782441	AQ782441 HS_3174_A
	8	87.6	9.1	839	122	CNS0122R	AL101037 Drosophil
c	9	87	9.0	1042	123	CNS0148K	AL103838 Drosophil
c	10	86.8	9.0	1198	120	B08337	B08337 T19F9-Sp6.1
c	11	85.8	8.9	870	101	AQ330286	AQ330286 nbxb0046J
	12	85.6	8.9	817	122	CNS009FM	AL053514 Drosophil
c	13	85.4	8.8	736	122	CNS009DE	AL053636 Drosophil
c	14	85.2	8.8	982	101	AQ325799	AQ325799 nbxb0021B
	15	84.8	8.8	1101	123	CNS0153V	AL104965 Drosophil
c	16	84.6	8.7	1223	120	B12981	B12981 T24D11-Sp6
	17	84.4	8.7	791	122	CNS009KS	AL053801 Drosophil
	18	83.8	8.7	822	114	AQ752069	AQ752069 HS_5570_B
	19	83.8	8.7	833	122	CNS007X3	AL050945 Drosophil
	20	83.6	8.6	710	71	AW349204	AW349204 GM210004A
	21	83.2	8.6	1046	122	CNS002KO	AL097794 Drosophil
c	22	83.2	8.6	1059	122	CNS002B	AL097133 Drosophil
c	23	82.8	8.6	732	96	AQ257374	AQ257374 nbxb0018K
c	24	82.4	8.5	569	101	AQ329762	AQ329762 nbxb0045P
c	25	82.4	8.5	956	101	AQ330169	AQ330169 nbxb0046L
c	26	81.8	8.5	840	96	AQ288571	AQ288571 nbxb0033I
	27	81.8	8.5	858	122	CNS0127J	AL101209 Drosophil
	28	81.6	8.4	507	91	W82081	W82081 me96h06.r1
c	29	81.6	8.4	870	116	AQ866797	AQ866797 nbhe0029E
c	30	81.4	8.4	865	96	AQ324474	AQ324474 mgxb0018B
c	31	81.4	8.4	1044	122	CNS00K3G	AL077176 Drosophil
c	32	81.2	8.4	864	64	AW155256	AW155256 mgie0002P
	33	80.8	8.4	830	122	CNS0118J	AL099949 Drosophil
c	34	80.6	8.3	506	122	CNS009K4	AL053777 Drosophil
c	35	80.6	8.3	844	120	B10796	B10796 T26G15-Sp6
c	36	80.6	8.3	1101	122	CNS000SX	AL050813 Drosophil
c	37	80.6	8.3	1101	122	CNS00215	AL097091 Drosophil
c	38	80.2	8.3	1147	120	B13042	B13042 T30M24-Sp6
c	39	80	8.3	949	101	AQ325830	AQ325830 nbxb0021F
c	40	79.6	8.2	700	46	A1906328	A1906328 PM-BT107-
c	41	79.6	8.2	815	116	AQ853920	AQ853920 nbxb0046G
c	42	79.6	8.2	968	113	AQ687544	AQ687544 nbxb0075I
	43	79.2	8.2	776	122	CNS009BD	AL053563 Drosophil
	44	78.8	8.1	952	117	AQ897460	AQ897460 HS_3134_A
c	45	78.6	8.1	833	102	AQ446640	AQ446640 nbxb0070F

ALIGNMENTS

RESULT 1
CNS00880
LOCUS CNS00880 593 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR16J23 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.
ACCESSION AL051540
VERSION AL051540.1 GI:4933381
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 593)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source Location/Qualifiers
1. 593
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR16J23"
/note="end : TET3"
BASE COUNT 448 a 25 c 18 g 10 t 92 others
ORIGIN

Query Match 9.9%; Score 95.8; DB 122; Length 593;
Best Local Similarity 41.9%; Pred. No. 2.9e-13;
Matches 224; Conservative 58; Mismatches 250; Indels 3; Gaps 1;

Qy 132 acaaacacacctcatcagagctgccacattggttcacaaatcagaaagacgagagtc 191
| | | | | : | | : | | : | | | | | | | | | |
Db 39 AAAATAAAACMGGAAAAAAACMAAAAAAAMAAAAAAMAAAAAAMAAAAAAMA 98
Qy 192 tgaatacaaacagccaaatcatcgaagagtacccaaacatgagaagcctgaaatga 251
| | | : : : | : : | | | | | | | | | | | |
Db 99 AMATAACMAAMMMACMAAAMAAAAAAMAGGACAAAAAAMMAAAMAAAAAAMA 158
Qy 252 caaggaggaaaaacacacccctgcaaacatcatcgaagagtacccagctcagcgaatc 311
| | | | | | | | | : | | | | | : | | | | | : |
Db 159 AAAACAAAAAAMAAAAAAMACRRAAAAAAMAAAAAAMAAAAAAMAAAAAAMA 218
Qy 312 gaaggagcagaaagagtagcagataaagaaaaacccgattcccccaatgggaaagccta 371
: : | | | | | | | | | | | | | : | | | | : |
Db 219 AMMAAAAAAAMAAAAAAMAAAAAAMAGGAGACMAAAMAAAAAAMAAAAAAMMA 278
Qy 372 agagcagcagaaacacgaagtcgaatccgaaatcccgagtacagagacaacaaga 431
| | : | | | : | : | : : | : : | | | | | | : |
Db 279 AAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 338
Qy 432 tgagaataagaaacataaagtagaagagtgccaggagtcacacgaatcgaagagcagca 491
| | | | : | | | : | | | | | | | | | : | | | : |
Db 339 MAAAAAAMRAAAMGAAAAAGGMAAAAAAAMAAAAAAMAAAAAAMAAAAAAMAA 398
Qy 492 agagtagcagaaagaaaaacccgattcccccaatgggaaagcctaagggcagcagaa 551
| | | | | | | | | : | : | : | : | | | | | | | |
Db 399 AAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAA 458

PT DNA construct contg. gene of interest controlled by cotton fibre
PT transcriptional factor - used to produce altered phenotype cotton
PT fibre cells expressing genes affecting pigmentation
PS Example 5; Fig 3A-J; 95pp; English.
CC The present sequence is a 4-4 cotton fibre expression cassette (version
CC II) from promoter construct pCGN5610. The lambda genomic phage clone used
CC to form this construct was designated 4-4(6). DNA constructs containing
CC cotton fibre-specific transcriptional factor promoters are useful to
CC produce cotton fibre cells with altered phenotype, especially altered
CC colour. Genes involved in the production of melanin (e.g. tyrosinase
CC gene and ORF438 encoded protein from Streptomyces antibioticus) and
CC indigo (mono-oxygenase genes possibly in conjunction with a
CC tryptophanase gene) are of interest. The promoters of the invention are
CC reliable and permit expression of a protein selectively in cotton fibre
CC to affect qualities such as fibre strength, length, colour and dyability
CC as required. The construct and methods can also be used for the
CC introduction of other advantageous genes into a cotton plant, e.g. a
CC plant hormone. In particular, fibres from a plant producing coloured
CC fibres may be used to produce yarns and/or fabrics that do not require
CC dyeing.
SQ Sequence 5518 BP; 1886 A; 794 C; 815 G; 2022 T;

Query Match 28.5%; Score 276; DB 1; Length 5518;
Best Local Similarity 86.6%; Pred. No. 3e-60;
Matches 322; Conservative 0; Mismatches 35; Indels 15; Gaps 1;

Qy 1 ctttctatttggttaaccatggctcatactttctgtcatctttcttcttccaaactt 60
Db 4139 CTTTCTATTGTTGATTAACCATGGCTCATAGCATTCGTCACCTTTCTCTCTTTCCAACTT 4198
Qy 61 ttactcattactgtctcactaatgatcggtagccacacgctgtctgcagcggtcgacat 120
Db 4199 TTACTCATAGTGTCTCACTAGTGACCGGTAGCCACTGTTTCGGCAGCGGCTCGACGT 4258
Qy 121 ttattccacacacacaacacctcatcagagctgccacaattggtctcaaaatagcaaaag 180
Db 4259 TTATTTCGAGACACAAGCAACCTCATCAGAGCTCCCAACAATGGCTTCAAAATACGAAAAG 4318
Qy 181 caggaagagtct-----gaatacaaacagccaaaatcatgaagagtac 225
Db 4319 CACGAAGAGTCTGAATACGAAAAGCCAGAAATACAAACAGCCAAAGATACGAGAGTAC 4378
Qy 226 ccaaaacatgagaagcctgaaatgtacaaggaggaagaaaaacacacacatcat 285
Db 4379 TCAAAACTTGAGAAGCCTGAAATGCAAAAGGAGGAAAAACAAAACCTGCAAAACAGCAT 4438
Qy 286 gaagagtaccacgagtcacgcgaatcgaaggagcagcaagagtacgataagaaaaacc 345
Db 4439 GAAGAGTACCACGAGTCACACGAATCAAGGAGCAGCAAAAGAGTACGAGAAAGAAAATCTC 4498
Qy 346 gatttccccaaa 357
Db 4499 GACGGGCCCGAA 4510

RESULT 11
T73865

ID T73865 standard; DNA; 5547 BP.
AC T73865;
DT 26-JAN-1998 (first entry)
DE Cotton fibre promoter clone 4-4(6) construct, pCGN5606 (Version I).
KW promoter; fibre-specific; transcriptional factor; promoter;
KW altered phenotype; colour; melanin; indigo; ss.
OS Gossypium hirsutum cv. coker 130.
FH Key Location/Qualifiers
FT misc_feature 1. 65
FT /*tag= a
FT /note= "fragment of pBluescriptII polylinker (as
FT stated in the specification)"
FT misc_feature 57. 5494
FT /*tag= b
FT /note= "genomic clone 4-4(6) from lambda phage clone of

a cotton Coker 130 genomic library (as stated in
the specification)"
FT misc_RNA 65. 4163
FT /*tag= c
FT /note= "5' flanking region of the 4-4(6) gene (as
FT stated in the specification)"
FT CDS 4163. 4502
FT /*tag= d
FT /note= "corresponds to part of the 4-4(6) ORF (as
FT stated in the specification)"
FT CDS complement (4131. 4502)
FT /*tag= i
FT /transl_except= (pos:4170. 4172, aa:Xaa)
FT /transl_except= (pos:4182. 4184, aa:Xaa)
FT /note= "Xaa = stop codon; No start or stop codons
FT given, possibly conforms to exon structure.
FT Encodes W21899"
FT misc_feature 4502. 4555
FT /*tag= e
FT /note= "synthetic polylinker oligonucleotide containing
FT unique target sites for EcoRI, SmaI, SalI, NheI
FT and BglII"
FT misc_feature 4163. 4555
FT /*tag= f
FT /note= "stuffer fragment left in place to facilitate the
FT monitoring of cloning manipulations (as stated in
FT the specification)"
FT 3'UTR 4555. 5494
FT /*tag= g
FT /note= "corresponds to the 940 nucleotides downstream of
FT the stop codon and constitutes the 3' flanking
FT region of the 4-4(6) gene (as stated in the
FT specification)"
FT misc_feature 5494. 5547
FT /*tag= h
FT /note= "fragment of pBluescriptII polylinker (as stated
FT in the specification)"
PN W09640924-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09897.
PR 07-JUN-1995; US-480178.
PR 01-JUL-1996; ZA-005572.
PA (CALJ) CALGENE INC.
PI McBride K, Pear JR, Perez-Grau L, Stalker DM;
DR WPI; 97-052325/05.
DR P-PSDB; W21899.
PT DNA construct contg. gene of interest controlled by cotton fibre
PT transcriptional factor - used to produce altered phenotype cotton
PT fibre cells expressing genes affecting pigmentation
PS Claim 22; Fig 2A-J; 95pp; English.
CC The present sequence is a 4-4 cotton fibre expression cassette (version
CC I) from promoter construct pCGN5606. The lambda genomic phage clone used
CC to form this construct was designated 4-4(6). DNA constructs containing
CC cotton fibre-specific transcriptional factor promoters are useful to
CC produce cotton fibre cells with altered phenotype, especially altered
CC colour. Genes involved in the production of melanin (e.g. tyrosinase
CC gene and ORF438 encoded protein from Streptomyces antibioticus) and
CC indigo (mono-oxygenase genes possibly in conjunction with a
CC tryptophanase gene) are of interest. The promoters of the invention are
CC reliable and permit expression of a protein selectively in cotton fibre
CC to affect qualities such as fibre strength, length, colour and dyability
CC as required. The construct and methods can also be used for the
CC introduction of other advantageous genes into a cotton plant, e.g. a
CC plant hormone. In particular, fibres from a plant producing coloured
CC fibres may be used to produce yarns and/or fabrics that do not require
CC dyeing.
SQ Sequence 5547 BP; 1889 A; 808 C; 822 G; 2028 T;

Query Match 27.7%; Score 268; DB 1; Length 5547;
Best Local Similarity 86.7%; Pred. No. 3.1e-58;
Matches 312; Conservative 0; Mismatches 35; Indels 13; Gaps 1;

Qy 527 gggaaaagcctaaagggcagcagaaacataaagccgaatatccgaaaatacctgagtgca 586
Db 470 GGGAAAAGCCTAAAGGGCAGCAGAAACATAAAGCCGATATCCGAAAATACCTGAGTGCA 529
Qy 587 agggaaaactagatgaggataaggaacataaacatgagttcccaagcatgaaaaagaag 646
Db 530 AGGAAAACCTAGATGAGGATAAGGAACATAAACATGAGTTCCCAAGCATGAAAAAGAAG 589
Qy 647 agggagaagaaacctgagaaagcagatgaccctgagtggttaaaatgcctgaatggccg 706
Db 590 AGGAGAAGAAACCTGAGAAAGCATAGTACCTGAGTGGGTAAATGCCTGAATGGCCG 649
Qy 707 aagtcacatgtttactcagtcgtgcgactaagccttaagccatatgacactggtgc 766
Db 650 AAGTCCATGTTTACTCAGTCTGGCTCGAGCACTAAGCCTTAAGCCATATGACACTGGTGC 709
Qy 767 atgtgccatcatcatgcagtaatttcaggatattgttaattattgttaataaaaaag 826
Db 710 ATGTGCCATCATCATGCAGTAATTCATGGGATATTGTAATTATATTGTAATAAAAAAG 769
Qy 827 atggtagtgaggaaatgtgtgtgcattcatcagc-agcaatgctgaatctctttgca 885
Db 770 ATGGTAGTGGGAAATGTGTGTGTCATCCATGTAGCAATGCTGAATCTCTTTGCA 829
Qy 886 tgcataagagattctgaatggttagttatgtttatctgtttctagtgaattaat 945
Db 830 TGCATAGAGATTCTGAATGTTATAGTTTATGTTATCGTTTCTAGTGAAATTAAT 889
Qy 946 tttgaatgttgatgtaattgtt 967
Db 890 TTTGAATGTTGATGTAATGTT 911

RESULT 7

T43360

ID T43360 standard; DNA; 3974 BP.
AC T43360;
DT 11-MAR-1997 (first entry)
DE Cotton FbLate2-82A gene and promoter.
KW FbLate; promoter; fibre; transgenic plant; cotton; ds.
OS Gossypium hirsutum var. Sea Island.
FH Key Location/Qualifiers
FT promoter 1. 2315
FT /*tag= a
FT /note= "the FbLate promoter located between
FT bases 1 and 2315 is preferred for use in
FT constructs of the invention"
FT cds 2315..3379
FT /*tag= b
FT /product= unidentified protein
PN W09639021-A1.
PD 12-DEC-1996.
PF 06-JUN-1996; U09449.
PR 06-JUN-1995; US-467504.
PA (MONS) MONSANTO CO.
PI John ME;
DR WPI; 97-042726/04.
PT Plant fibre-specific, developmentally regulated FbLate promoter -
PT useful for producing transgenic plants, esp. cotton, with altered
PT fibre properties
PS Claim 4; Page 57-59; 79pp; English.
CC A 3974 bp region (T43360) of clone pSKSIFbLate2-28A includes the
CC fibre-specific FbLate promoter that is active during late fibre
CC development, plus a coding sequence (FbLate-82A) for an unknown
CC protein. The clone was obtd. from a fibre genomic library using a
CC cDNA clone (see also T43362) that corresponds to RNA prevalent in
CC late fibre development, and insertion of an isolated clone into
CC Bluescript SK+ vector. The FbLate promoter can be used for tissue-
CC and developmental-specific expression of fibre and non-fibre
CC proteins (e.g. polyhydroxybutyrate biosynthetic enzymes) in
CC transgenic plants, esp. to alter the fibre characteristics of
CC cotton.
SQ Sequence 3974 BP; 1523 A; 603 C; 597 G; 1251 T;

Query Match 53.1%; Score 513.4; DB 1; Length 3974;
Best Local Similarity 85.9%; Pred. No. 7.1e-120;
Matches 587; Conservative 0; Mismatches 81; Indels 15; Gaps 1;

Qy 8 ttgtgtaaacatggtcctaactttctgcatcctttctccttttccaacttttactca 67
Db 2304 TTCGGTTAACCATGGCTCATAACACTCGTCACCTTTCTTCTTTTCCAACCTTTACTCA 2363
Qy 68 ttactgtctcactaatgatcggtagccacacgctctcgcagcggtcgacattttattcc 127
Db 2364 TTAGTGTCTCACTAATGATCGGTAGCCACACCTCTCGACAGCGGCTCGAGTTTATTCG 2423
Qy 128 acacacaaacacccatcagagctgccacaattggtctcaaaatacgaagcagcaag 187
Db 2424 AGACACAAACACCTCATCGGAGTTGCCACAATTAGTCTCAAAATACGAAAGCAGGAAG 2483
Qy 188 agtct-----gaatacaaacagccaaatatcatgaagagtacccaaac 232
Db 2484 AGTCTGAATATGAAAGCCGGAATACAAACAGCCAAAGTATGACGAGAGTACCCAAAC 2543
Qy 233 atgagaagcctgaatgtacaaggaggaacacacacacccctgcaacatcatgaagagt 292
Db 2544 ATGAGAAGCCTGAATTCACAGGAGGAAAAACAAACCGTGCAAGCAACATGAAGAGT- 2603
Qy 293 acccagagtcacggaatcgaaggagcagcaagagtacgataaagaaaaacccgatttcc 352
Db 2604 ACCCAGAGTCACACAAATCGAAGGAGCAGCAAGAGTACCAGAAAGAAACCCGAGTTCC- 2663
Qy 353 ccaaatgggaaaagcctaaagagcagcagaaacacgaagtcgaatatccgaaaatacccg 412
Db 2664 CCAAAATGGAAAAGCCTAAAGAGCAGGAGAAACACGAAGTCAATATCCGAAAATACTCG 2723
Qy 413 agtacaaggaacaaacagatgagaataagaacataaagatgaagagtgccaggagtcac 472
Db 2724 AGTACAAGGAAAACCAAGATGAGGGTAAGGAACATAAACATGAAGAGTACCAGAACTAC 2783
Qy 473 acgaatcgaagagcagcaagagtagcagagaagaaaaacccgatttcccaaatgggaaa 532
Db 2784 GTGAATCGAAGGAGCAGCAAGAGTACGAGAAAGAAAAACCCGAGTTCCCAAAATGGAAA 2843
Qy 533 agcctaaagggcagcagaaacataaagccgaatatccgaaaatacctgagtgcaaggaaa 592
Db 2844 AGCCTAAAGAGCAGCAGAAACACGAAGTCAATATCCGAAATACCCAGTACAGAGAA 2903
Qy 593 aactagatgaggataaggaacataaacatgagttcccaagcatgaaaaagaaggagaga 652
Db 2904 AGCAAGATGAGGGTAAGGAACATAAACATGAGGAGTGCCACAAGTACACGAATCGAAG 2963
Qy 653 agaaacctgagaaagggcatagta 675
Db 2964 AGCAGGAAGAGTACGAGAAAGAA 2986

RESULT 8

T43362

ID T43362 standard; cDNA; 645 BP.
AC T43362;
DT 11-MAR-1997 (first entry)
DE Cotton FbLate 2-82A gene cDNA clone A11 (FbLate-2).
KW FbLate; promoter; fibre; transgenic plant; cotton; ds.
OS Gossypium hirsutum.
PN W09639021-A1.
PD 12-DEC-1996.
PF 06-JUN-1996; U09449.
PR 06-JUN-1995; US-467504.
PA (MONS) MONSANTO CO.
PI John ME;
DR WPI; 97-042726/04.
PT Plant fibre-specific, developmentally regulated FbLate promoter -
PT useful for producing transgenic plants, esp. cotton, with altered
PT fibre properties

RESULT 4

T62624

ID T62624 standard; cDNA to mRNA; 1283 BP.
AC T62624;
DT 14-MAY-1997 (first entry)
DE Cotton fibre specific cDNA clone CKFB15-E9.
KW cotton; fibre-specific; strength; transgenic plant; anthesis;
KW developmentally regulated; E6; H6; antisense; sense; ss.
OS Gossypium hirsutum strain Coker 312.
PN US5597718-A.
PD 28-JAN-1997.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-OCT-1993; US-138814.
PR 20-SEP-1995; US-530797.
PA (CETU) AGRACETUS.
PI Brill WJ, John ME, Umbeck PF;
DR WPI; 97-108326/10.
PT Prodn. of transgenic cotton plants - by transformation with the H6
PT coding sequence or E6 anti-sense sequence, produces fibre of altered
PT strength
PS Example 4; Column 53-54; 33pp; English.
CC T62609-24 are cotton fibre-specific cDNA clones which can be used to
CC identify genomic clones. This clone, CKFB15-E9, is expressed in fibre
CC cells, but is also expressed at low levels in petal. (CK =
CC Coker; FB = Fibre; 10, 15 or 23 = age in days of fibre cells; A1 and the
CC last character and number stand for clone identity). The fibre-specific
CC genes were identified by differential cDNA library screenings. Coding
CC sequences from these isolated genes are used in sense or antisense
CC orientation to alter the fibre characteristics, e.g. strength, of
CC transgenic fibre-producing plants.
SQ Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

Query Match 75.5%; Score 730; DB 1; Length 1283;
Best Local Similarity 82.5%; Pred. No. 2.2e-174;
Matches 937; Conservative 0; Mismatches 30; Indels 169; Gaps 2;

Qy 1 ctttctatttggttaaccatggtcctataactttctgtcatcctttcttctttccaactt 60
Db 28 CTTTCTATTGTTAAACCATGGCTCATAACTTTTGTCATCCTTCTCTCTTTCCAACCT 87
Qy 61 ttactcattactgtctcactaatgatcggtatgccacacgtctctgtcagcggtcgacat 120
Db 88 TTACTCATTACTGTCTCACTAATAATCGGTAGTCACACCGTCTCGTCAGCGGCTCGACAT 147
Qy 121 ttattccacacacaaacacctcatcagagctgccaaattggtctcaaaatacgaag 180
Db 148 TTATTCCAGACACAAACACCTCATCAGAGCTGCCAATTTGGCTCAAAATACGAAAG 207
Qy 181 cacgaagagctctgaatacaaacagccaaatatcatgaagagtaccacaaacatgagaag 240
Db 208 CACAAGAGTCTGAATACAAACCAAAATATCAGAAAAAGTACCAAAACATGAGAAG 267
Qy 241 cctgaaatgtacaaggaggaaaaacaaacccctgcaaacatcatgaagagtaccacgag 300
Db 268 CCTAAATGCACAAGGAGGAAAAACAAACCCCTGCAAAACATCATGAAGAGTACCAGAG 327
Qy 301 tcacgcgaatcgaaggagcacgaagagtacgataagaaaaacccgatttccccaaatgg 360
Db 328 TCACGCGAATCGAAGGAGCAGGAAGTAGCATTAAGAAAAACCCGATTTCCTCAATGG 387
Qy 361 gaaaagcctaagagcagcagaaacaggaagtcgaatatccgaaaatacccgagtacaag 420
Db 388 GAAAGCCCTAAGAGCACAAGAACACGAAGTTGAATATCCGAAAATACCGAGTACAG 447
Qy 421 gacaacaagatg----- 433
Db 448 GACAACAAGATGAGGATAAGGAACATAAAATGAAGAGTACCATGAATCAGCGAATCG 507

Qy 434 ----- 433
Db 508 AAGGAGCACGAATAACGAGAAAGAAAAACCCGAGTTCCTCCAAATGGGAAAGCCTAAA 567
Qy 434 ----- 433
Db 568 GAGCAGGAGAAACACGAAGTGAATATCCGAAAATACCCGAGTACAAGGAAAGCAAGAT 627
Qy 434 -agaataagaacataaagatgaagagtgccaggagtcacacgaatcgaagagcagcaa 492
Db 628 AAGAGTAAGGAACATAAAGATGAAGAGTGCACGAGTGCACGAATCGAAAGATCAGCAA 687
Qy 493 gagtacgagaaagaaaaacccgatttccccaaatgggaaaacctaagggcagcagaaa 552
Db 688 GAGTACGAGAAAGAAAAACCAATTTCCTCAATGGGAAAGCCTAAGAGCAGCAGAGAA 747
Qy 553 cataaagccgaatatccgaaaatacctgagtgcaggaaaaactagatgaggataaggaa 612
Db 748 CATAAGCCGAATATCCAAAAATACCCGAGTGAAGGAAAAACAAGATGAGGATAAGGAA 807
Qy 613 cataaacatgagttcccaagcatgaaaaagagagagagaaacctgagaaaggcata 672
Db 808 GATAAACATGAGTTCCTCAAGCATGAAAAAGAGAGAGAGAAACCTGAGAAAGGCAGA 867
Qy 673 gtaccctgagtggttaaaatgctgaatggcgaagtcacatgtttactcagtcgtgctc 732
Db 868 GTACCTGAGTGGGTTAAATGCTGAATGGCCGAGTCCATGTTTACTCAGTGTGCTC 927
Qy 733 gacactaagccttaagccatgatgacactggtgcatgtgccatcatcagtaatttc 792
Db 928 GAGCATTAGCCTTAAGCCATATGACACTGGTGCATGTGCCATCATCATGAGTAATTC 987
Qy 793 atgggatattgtaattattgtttaataaaaaagatggtagtggaatgtgtgtgtgc 852
Db 988 ATGGGATATCGTAATTATATTGTTAATAAAAAAGATGGTGAAGTGGGAAATGTGTGTGC 1047
Qy 853 attcatccatg-agcaatgctgaatctctttgcatgcatagagattctgaatgttatag 911
Db 1048 ATTCATCCATGTAGCAATGCTGAATCTCTTTGCATGATAGAGATTCTGAATGGTTATAG 1107
Qy 912 tttatgttatatcgtttcttagtgaaattaattttgaatgtgtatgtatgtt 967
Db 1108 TTTATGTTATATCGTTTGTCTAGTGAAATTAATTTGAATGTGTATCTAATGTT 1163

RESULT 5

T70055

ID T70055 standard; cDNA; 1283 BP.
AC T70055;
DT 20-AUG-1997 (first entry)
DE Cotton fibre specific cDNA clone E9.
KW cotton; E6; fibre; promoter; transgenic plant; truncated;
KW heterologous gene expression; ds.
OS Gossypium hirsutum strain Coker 312.
PN US5620882-A.
PD 15-APR-1997.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-MAY-1992; US-885970.
PR 19-OCT-1994; US-298829.
PA (CETU) AGRACETUS INC.
PI John M;
DR WPI; 97-235185/21.
PT DNA constructs contg. truncated promoter sequence - for
PT fibre-specific gene expression in cotton plants
PS Example 3; Column 45-48; 48pp; English.
CC T70040-57 are cotton fibre-specific cDNA clones which can be used to
CC obtain genomic clones containing fibre-specific promoters. Claimed DNA
CC constructs comprise a truncated promoter sequence (from one of T70031-38)
CC that promotes preferential gene expression in plant fibre cells, a
CC protein coding sequence not naturally associated with the promoter
CC sequence and a 3' termination sequence. The DNA constructs are useful for

Db 121 TTATTCCACACACAAACCTCATCAGAGCTGCCAATTTGGCTTCAAAATACGAAAG 180
Qy 181 cacgaagagctctgaatacaaacagccaaatatcatgaagagatcccaaacatgagaag 240
Db 181 CACGAAGAGTCTGAATACAAACAGCCAAATATCATGAAGAGTACCCAAACATGAGAAG 240
Qy 241 cctgaatgtacaaggaggaaaaaaccctgcaaacatcatgaagagatccacagag 300
Db 241 CCTGAATGTACAGGAGGAAAAACCAACCTGCAAAATCATGAAGAGTACCACGAG 300
Qy 301 tcacgcgaatcgaaggagcagcaagagtagcagataaagaaaaaccgattccccaaatgg 360
Db 301 TCACGCGAATCGAAGGAGCAGCAAGAGTAGCATAAGAAAAACCGATTTCGCCAAATGG 360
Qy 361 gaaaagcctaagagcagcagaaacacgaagtcgaatatccgaaatacccgagtacaag 420
Db 361 GAAAAGCCTAAGAGCAGGAGAACACGAAGTCGATATCCGAAATACCCGAGTACAG 420
Qy 421 gacaacaagatgagaataaagaaacataaagatgaagagtgccaggagtcacacgaatcg 480
Db 421 GACAAACAGATGAGAATAAGAAACATAAAGATGAAGAGTGCCAGGAGTCACACGAATCG 480
Qy 481 aaagagcagcaagagtagcagaaagaaaaaccgattccccaaatgggaaaagcctaaa 540
Db 481 AAAGAGCAGCAAGAGTAGCAGAAAGAAAAACCGATTTCGCCAAATGGGAAAAGCCTAAA 540
Qy 541 gggcagcagaaacataaagccgaatatccgaaatacctgagtgcaaggaaaactagat 600
Db 541 GGGCAGCAGAAACATAAAGCCGAATATCCGAAATACCTGAGTCAAGGAAAAATAGAT 600
Qy 601 gaggataaggaacataaacatgagttcccaagcatgaaaaagaaggagaagaaacct 660
Db 601 GAGGATAAGGAACATAAACATGAGTTCCCAAGCATGAAAAAGAGAGAGAAGAAACCT 660
Qy 661 gagaaggcatagtagcctgagtggttaaatgcctgaatggccgaagtcattgtttac 720
Db 661 GAGAAAGGCATAGTACCTGAGTGGTTAAATGCCTGAATGGCCGAAGTCCATGTTTAC 720
Qy 721 tcagtcgtgctgcagcactaagccttaagccatgacactggtgcatgtgccatcatca 780
Db 721 TCAGTCTGGCTCGAGCACTAAGCCTTAAGCCATATGACACTGGTGCATGTGCCATCATCA 780
Qy 781 tgcagtaatttcattggtatattgttaataaaaaagatggtgagtgaggaa 840
Db 781 TGCAGTAATTTTATGGGATATTGTAATTATATTGTTAATAAAAAAGATGGTGAAGTGGAA 840
Qy 841 atgtgtgtgtgcattcatccatgagcaatgctgaatctcttgcagcatagagattctg 900
Db 841 ATGTGTGTGTGCATTCATCCATGAGCAATGCTGAATCTCTTTGCATGCATAGAGATTCTG 900
Qy 901 aatggttatagtttatgttatatcgttgttctagtaaatattttgaatgtgtatg 960
Db 901 AATGTTATAGTTTATGTTATATCGTTTGTCTAGTGAATTAATTTGAATGTGTATG 960
Qy 961 taatgtt 967
Db 961 TAATGTT 967

RESULT 2

T13048

ID T13048 standard; cDNA; 1283 BP.

AC T13048;

DT 27-MAY-1996 (first entry)

DE Cotton fibre-specific cDNA clone E9.

KW Cotton; fibre; promoter; transgenic plant; crop improvement; ds.

OS Gossypium hirsutum strain Coker 312.

PN US5495070-A.

PD 27-FEB-1996.

PF 04-OCT-1988; 253243.

PR 04-OCT-1988; US-253243.

PR 21-NOV-1990; US-617239.

PR 18-MAY-1992; US-885970.

PA (CBTU) AGRACETUS INC.

PI John M;

DR WPI; 96-139095/14.

PT New isolated fibre-specific promoters - used for introducing

PT altered fibre-specific characteristics into plants, partic. cotton.

PS Example 3; Column 45-46; 48pp; English.

CC Cotton cDNA clone E9 (T13048) was isolated from a cDNA library of

CC cotton var. Coker 312 15-day-old boll cells using a subtractive

CC hybridization procedure. The clone hybridises strongly to fiber

CC RNA and weakly to petal RA. E9 and other fibre-specific cDNA clones

CC (see T13033-47 and T13049-T13050) were used to screen cotton genomic

CC libraries, leading to the isolation of genomic clones (see T13025-32

CC and T13052-53) contg. sequences capable of promoting gene expression

CC

SQ Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

Query Match 75.5%; Score 730; DB 1; Length 1283;

Best Local Similarity 82.5%; Pred. No. 2.2e-174;

Matches 937; Conservative 0; Mismatches 30; Indels 169; Gaps 2;

Qy 1 ctttctatttggtaaccatggctcataactttgtcatcctttcttcttccaaactt 60
Db 28 CTTTCTATTTTGTAAACCATGGCTCATAACTTTTGTCTCCTTTCTCTTTTCCAACTT 87
Qy 61 ttactcattactgtctcactaatgatcggtagccacacgctctcgtcagcggtcgacat 120
Db 88 TTACTCATTACTGTCTCATAATAATCGGTAGTCACACCGTCTCGTCAGCGGCTCGACAT 147
Qy 121 ttattccacacacaaacacctcatcagagctgccacaattggcttcaaaatcagaaaag 180
Db 148 TTATTCCAGACACAAACACCTCATCAGAGCTGCCAATTTGGCTTCAAAATACGAAAG 207
Qy 181 cacgaagagctctgaatacaaacagccaaatatcatgaagagatcccaaacatgagaag 240
Db 208 CACAAGAGTCTGAATACAAACAAACAAATATCAGAAAAGTACCCAAACATGAGAAG 267
Qy 241 cctgaatgtacaaggaggaaaaaaccctgcaaacatcatgaagagatccacagag 300
Db 268 CCTAAAATGCACAGGAGGAAAAACCAACCTGCAAAATCATGAAGAGTACCACGAG 327
Qy 301 tcacgcgaatcgaaggagcagcaagagtagcagataaagaaaaaccgattccccaaatgg 360
Db 328 TCACGCGAATCGAAGGAGCAGCAAGAGTAGCATAAGAAAAACCGATTTCGCCAAATGG 387
Qy 361 gaaaagcctaagagcagcagaaacacgaagtcgaatatccgaaatacccgagtacaag 420
Db 388 GAAAAGCCTAAGAGCACAAGAAACACGAAGTTGAATATCCGAAATACCCGAGTACAG 447
Qy 421 gacaacaagatg----- 433
Db 448 GACAACAGATGAGGATAAGGAACATAAAATGAAGAGTACCATGAATCAGCGAATCG 507
Qy 434 ----- 433
Db 508 AAGGAGCACGAAGATACGAGAAAGAAAAACCGAGTTCGCCAAATGGGAAAGCCTAAA 567
Qy 434 ----- 433
Db 568 GAGCAGAGAAACACGAAGTCGAATATCCGAAATACCCGAGTACAAGGAAAGCAAGAT 627
Qy 434 -agaataagaacataaagatgaagagtgccaggagtcacacgaatcgaagagcacgaa 492
Db 628 AAGAGTAAGGAACATAAAGATGAAGAGTGCCACGAGTCACACGAATCGAAGATCACGAA 687
Qy 493 gagtacgagaaagaaaaaccgattccccaaatgggaaaagcctaaggagcagagaaa 552
Db 688 GAGTACGAGAAAGAAAAACCAATTTCTCAANTGGGAAAGCCTAAGAGCACGAGAAA 747
Qy 553 cataaagccgaatatccgaaatacctgagtgcaaggaaaaactagataggataaggaa 612
Db 748 CATAAAGCCGAATATCCAAATAACCCGAGTCAAGGAAAAACAAGATGAGGATAAGGAA 807

```

      ||| | | | ||| ||| | ||| ||| ||| ||| | | |
Db 3514 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3455
Qy 288 agagtaccacgagtcacgcgaatcgaaggagcacgaagagtcagataaagaaaaaccga 347
      ||| | | | ||| ||| | ||| ||| ||| ||| | | |
Db 3454 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3395
Qy 348 tttcccaaatgggaaagcctaagagcagcagaaacacgaagtgcgaatccgaaat 407
      | | | | ||| | | | | | | | | | | | |
Db 3394 AAAAAAAAAAAAAAAAAANNNNAANNNAAAAAAAAAAAAAAAAANNNNAANANANAAAAAAAA 3335
Qy 408 acccgagtacaaggacaaacagatgagataagaacataaagatgaagagtgccagga 467
      | | | | | | | | | | | | | | | | | |
Db 3334 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3275
Qy 468 gtcacacgaatcgaagagcagcgaagagtcagagaagaaaaaccgatttcccaaatg 527
      | | | | ||| | | | | | | | | | | | |
Db 3274 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3215
Qy 528 ggaaagcctaagggcagcagaaacataaagccgaatatccgaaatccctgagtgcaa 587
      ||| | | | ||| | | | ||| | | | | | |
Db 3214 AAAAAAAAAANNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNA 3155
Qy 588 ggaaaactagatgagataaggaacataaacatgagttcccaagcatgaaaaagaaga 647
      ||| | | | ||| | | | ||| | | | | | |
Db 3154 AAAAAAAAAANNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANNA 3095
Qy 648 ggagaagaaacctgagaagcagatagtcacctgagtggttaaatgcctgaatggccga 707
      ||| ||| | | | | | | | | | | | | | |
Db 3094 AAAAAAAAAACAAAAANNNNAANNNNANNNNANNNNAGNNNNATTATNACCCC 3035
Qy 708 agtccatgtttactcagtcggtcgagcactaagccttaagccatat 755
      | | | ||| ||| | | | | | | | | | |
Db 3034 CCCCTGGTATGATTGCATCTTGCTCANACCAAAAAAGTTATTAATAT 2987

```

Search completed: September 3, 2000, 02:53:40
Job time: 31570 sec

Qy 168 aaatcatgaaagcagaagagtctgaatacaaacacgcaaaaatatcatgaagagtacc 227
|| || | | | | | | | | | | | |
Db 2559 AACACATGAAGAAGAAAAAAGTAACATATGAAGAAGAAGAAGAAGAAAAGT 2618

Qy 228 aaaacatgagaagcctgaagtgtacaaggagggaaaaacaaacctgcacaacatcatga 287
| | | | | | | | | | | | | | | |
Db 2619 AACACATGAAGAAGAAAAAAGTAACACATGAAGAAGAAGAAAAGTAAACACATGAAGA 2678

Qy 288 agagtaccacgagtcacgcgaatcgaaaggagcagaagagtagcataaagaaaaaccga 347
|| | | | | | | | | | | | | | |
Db 2679 AGAAGAAAAGTAAATACATGAAGAAGAAGAAAAGAAGAAGATGAGGAAGAAGAAGAAGA 2738

Qy 348 ttcccccaaatgggaaaagccctaagagcacgagaaacacgaagtcgaatatccgaaat 407
| | | | | | | | | | | | | | | |
Db 2739 AGAAGAAGAAGAAGAGGAAGAAGAAGAAGAAGATGAGGAAGAAGAAGAAGAAGAAGA 2798

Qy 408 acccgagtacaaggacaaacaagatgagaataagaacataaagatgaagagtgccagga 467
| | | | | | | | | | | | | | | |
Db 2799 AGAAGATGAGGAAGAGGAAGAAGAAGAAAATGAGGAAGAAGAAGAAGAAGAAATAA 2858

Qy 468 gtcacacgaatcgaaagagcacgaagagtagcagaagaaaaaccgatttcccccaaatg 527
| | | | | | | | | | | | | | | |
Db 2859 GGAAGAAGAAGAAGAAAAAAGAGCATGAAGAAGAAGTAAACATGAAGAAGAAGA 2918

Qy 528 ggaaaaagcctaaggggcagagaaacataaagccgaatatccgaaaaacttgagtgcga 587
| | | | | | | | | | | | | | | |
Db 2919 AGAAAAAGTAACACATGAAGAAGAAAAAAGTAACACATGAAGAAGAAGAAATGTAA 2978

Qy 588 ggaaaaactagatgaggataaggaacataaacatgagttcccaagcatgaaaaagaaga 647

```

RESULT 15
AC013349/c
LOCUS AC013349 129404 bp DNA HTG 06-FEB-2000
DEFINITION Homo sapiens clone RP11-22K1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC013349
VERSION AC013349.2 GI:6910730
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 129404)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-22K1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 129404)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 6, 2000 this sequence version replaced gi:6272406.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4134
Center clone name: 22_K_1
-----
* NOTE: This record contains 151 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 930: contig of 930 bp in length
* gap of unknown length
*
* 931 1855: contig of 925 bp in length
* gap of unknown length
*
* 1856 2759: contig of 904 bp in length
* gap of unknown length
*
* 2760 3706: contig of 947 bp in length
* gap of unknown length
*
* 3707 4566: contig of 860 bp in length

```

gene
CDS

```
/db_xref="GI:1255419"
/translation="MLCQVCQAGPEPHFGGISCRACAAFFRRYVHSRKLDISCTCKH
RLATSHPCRCLMKCMATGMVCKVQGSREKNITTSPLPGHISSISLARSIVPRD
CSNISCTSVSKWTVEKMRKLDYGEKICEINFTQFSSFVKRDTLLWDLGEKIFTDVKL
LSEADKHSILCNFFPRWMLDSAVAICPDYEEKSAYIKSKDYDMLLHFTGSSMPKEK
RLKDHEILKIFKPYWDFPHYETAVPIHFKKLDKIEYMAIFLLLLPDDAYTNISEEGVK
LCQNVKRVVQRELKGYQTDSCNDEMRFVETMDTLLLEKAEKIQEEVLICGFNNVTL
HEDFTIIFQVKKL"
complement(15326..17135)
/gene="C33G8.11"

CDS
complement(join(15326..15621,15852..16191,16239..16466,
16545..16636,16682..16747,16994..17135))
/gene="C33G8.11"
/note="similar to steroid/thyroid/retinoic nuclear hormone
receptors"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC25855.1"
/db_xref="GI:3294498"
/translation="MVNCMVCDASSAQVHFGAICRACAAFFRRYVNSKKLTILCKCL
SKKESQYPCRLCKRMKCKKAVGMEASKVQGPDRVNNPFRKIMIEGSSSPSTLSSIEYG
LQPROSELIKLIKNYNKLEMTREAIYNMPTSTGTGMQVNLVELSEVKTDSKLMWKL
CEDTFPEFNPQLCKMDKRIFLNNFYKSWGILEVTMLATKYNDSKNFYSPSGAMCTSI
NE FYVNTVRDNSAISQEDIIRYLATRVIVFINLWLRVFEPPYWKYHFQEVIDPFLALKLNDM
ENMALPGIIFWDSYTNISDELSLCHSMRKII CRELSAHNPETCTTSSRFFETDITL
NIEEVSAMTTRSQCCKLLQLKAEKRCQBEIALCGFYNFVEDDDMKNMIMWEKY"
18885..20672
/gene="nhr-42"

CDS
join(18885..18967,19407..19725,19771..19896,19988..20051,
20100..20211,20260..20375,20422..20672)
/gene="nhr-42"
/note="C33G8.6; coded for by C. elegans cDNA yk482f11.5;
coded for by C. elegans cDNA yk482f11.3"
/codon_start=1
/product="similar to steroid/thyroid/retinoic nuclear
hormone receptors"
/protein_id="AAC25857.1"
/db_xref="GI:3294500"
/translation="MTRQTTTSQTCCLICGSDASLHFGALSCRACAAFFRRKRVAGRRNI
FRRCRQCKVDTGMRKLKASCRCYDKCLKLVGMRESAVLSRLAKKNQNYKKSIVGSPDAY
EPSTSTSDSVLENLOSAYHKLLETRKRVFNISETHVSOCCKYKRNMDVFFEDIKLWME
```

Query Match 9.2%; Score 89; DB 34; Length 39103;
Best Local Similarity 48.7%; Pred. No. 6.2e-09;
Matches 242; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

Qy 169 aaatacgaaaagcacgaagagtctgaaatacaaacagccaaaaatatcatgaagagtaccca 228
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31431 AAGGAAGAAGATGATCAGGACGGAAAGCAAAAAAGATAAGAAAAAAAAGACGAGAGAAT 31490

Qy 229 aaacatgagaagcctgaagtgtacaaggaggaaaaaacaaaaccttgcaaacatcatgaa 288
| | | | | | | | | | | | | | | | | |
Db 31491 GACGATGATGATGAAGAAGACAAGAAGAAGACAAGAAGAAAAAGAGGATGATAATGAT 31550

Qy 289 gagtaccacgagtcacgcgaatcgaaggagcacgaagtacgataaagaaaaaccgat 348
||| | | | | | | | |
Db 31551 GAGGAGGAAAAAGAGAAGATAAGAAAAAAGGACAAAGAAAAGGCATGATGACCAT 31610

Qy 349 tcccccaatgggaaagcctaagaagacagagaaacacgaagtgcgaatatccgaaaata 408
|| ||| || | || | || | || |
Db 31611 GAAGATGAAATGATAAGAAAAAGGATAAGAAAAAAGGAAGTACAAAGGTATGATGAT 31670

Qy 409 cccgagtacaaggacaaacaagatgagataagaancataaagatgaagagtccaggag 468
| | | | | | | | | | | | | | | | | |
Db 31671 GAGAATGAGGATGACAAAAAAGGAAGCAAAAAGAAAGAGGATGATAAGGAAAAGGAT 31730

Oy 469 tcacacgaatcgaagaagcacgaaggtacgagaagaaaaaccggatttccccaatatgg 528
| | | | | | | | | | | | | | | | | |
Db 31731 GATGATGATGAGGAAGAGAAGGATAAGAAAAAAGACAAAAAGAAGTACCACGC 31790

Oy 529 gaaaagcctaaggcgacgagaacaataaagccgaatatccgaaataccctgagtgcag 588
||| ||| ||| ||| ||| ||| ||| |||
Db 31791 GAAGAAGTATAAAAAAGGACAAAGAAAAGAAAAGGACGTATGCGTAGGAGGTAG 31850

Qy 589 gaaaaactagatgaggataaggaacataaacatgagttcccaagcatgaaaaagaag 648
||| | ||||||| | | | | | | ||| |
Db 31851 GATAAATAGAAAAAGGATAACAAAAAGAGAAGATGATAAGGTATGAAGATGAGGA 31910

Qy 649 gagaagaaacctgagaa 665
 || ||| |||
 Db 31911 AAGGAGAAGGACAAGAA 31927

```

RESULT 12
AC006884/c
LOCUS      AC006884 193188 bp DNA HTG 26-FEB-1999
DEFINITION Caenorhabditis elegans clone Y57E12, *** SEQUENCING IN PROGRESS
            ***; 4 unordered pieces.
ACCESSION  AC006884
VERSION    AC006884.2 GI:4309911
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (bases 1 to 193188)
AUTHORS    Waterston,R.H.
TITLE      The sequence of Caenorhabditis elegans clone
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 193188)
AUTHORS    Waterston,R.H.
TITLE      Direct Submission
JOURNAL    Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT    On Mar 1, 1999 this sequence version replaced qi:4263464.

```

```

COMMENT      On Mar 1, 1999 this sequence version replaced gi:263464.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1      2784: contig of 2784 bp in length
*      2785      2799: gap of unknown length
*      2800      37804: contig of 35005 bp in length
*      37805      37819: gap of unknown length
*      37820      74165: contig of 36346 bp in length
*      74166      74180: gap of unknown length
*      74181      193188: contig of 119008 bp in length.

```

```

FEATURES             Location/Qualifiers
    source            1..193188
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y57E12"
BASE COUNT   63610 a 33574 c 33884 g 62075 t      45 others
ORIGIN

```

Query Match 9.2%; Score 89; DB 41; Length 193188;
Best Local Similarity 48.7%; Pred. No. 5e-09;
Matches 242; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

Qy 169 aaatacgtaaaagcacgaagagtctgaatacaaacagccaaatatcatgaagagtaccca 228
|| ||| ||| | | | | | | | | | |
Db 125873 AAGGAAGAAGATGATGAGGAGGAGAGCAAAAAAGATAAGAAAAAAAAAAGCAGAGAAGAAAT 125814

Qy 229 aaacatgaggaagcctgaaatgtacaaggaggaaaaacaaaacacctgcaaacatcatgaa 288
| | | | | | | | | | | | | | | | | | | |
Db 125813 GACGATGATGATGAAGAAGACAAGAAAGACAGAGAAGAAAAAGAGGATGATAATGAT 125754

Oy 289 gattaccacgagtcacgcgaatcgaaggcacgcaagagtacgataaaagaaaaaacccgat 348
||| || |
Db 125753 GAGGACGAAAAAAGAAGATGAAGAAAAGAAGGCACAAGAAAGGCCGTGATGCAGT 125694

```
/rpt_family="ORR1A3
```

```

RESULT 11
CEL33G8
LOCUS      CEL33G8      39103 bp      DNA            INV            08-JUL-1998
DEFINITION Caenorhabditis elegans cosmid C33G8.
ACCESSION  U53154
VERSION    U53154.1   GI:1255414
KEYWORDS
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (bases 1 to 39103)
AUTHORS    Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,

```

```
repeat_region 21957..22257
/rpt_family="AluSx"
repeat_region complement(22438..22538)
/rpt_family="MLT1J"
repeat_region complement(23375..23667)
/rpt_family="AluY"
repeat_region 24553..24858
/rpt_family="AluY"
repeat_region complement(25021..25234)
/rpt_family="MIR"
repeat_region 26957..27261
/rpt_family="AluY"
repeat_region complement(27524..27669)
/rpt_family="MLT1B"
repeat_region complement(27670..27964)
/rpt_family="AluSp"
repeat_region complement(27965..28210)
/rpt_family="MLT1B"
repeat_region complement(28228..28522)
/rpt_family="AluSg"
repeat_region complement(28592..28700)
/rpt_family="MIR"
repeat_region complement(29353..29537)
/rpt_family="MIR"
repeat_region 29802..30095
/rpt_family="AluSg"
repeat_region complement(30099..30219)
/rpt_family="MIR"
repeat_region 30227..30267
/rpt_family="(TG)n"
repeat_region 30276..30321
/rpt_family="(CA)n"
repeat_region complement(30322..30611)
/rpt_family="AluJo"
repeat_region complement(30956..31151)
/rpt_family="MER3"
repeat_region complement(31533..31633)
/rpt_family="L2"
repeat_region 31810..31858
/rpt_family="L2"
repeat_region 32028..32107
/rpt_family="L1PA4"
repeat_region complement(32919..33182)
/rpt_family="MIR"
repeat_region 34169..34314
/rpt_family="CT-rich"
repeat_region 34774..35007
/rpt_family="MIR"
repeat_region 35589..35621
/rpt_family="AT-rich"
repeat_region 35763..36142
/rpt_family="MER41A"
repeat_region 36143..36452
/rpt_family="AluSg"
repeat_region 36453..36636
/rpt_family="MER41A"
repeat_region 38011..38071
/rpt_family="CT-rich"
repeat_region 38083..38262
/rpt_family="(TTC)n"
repeat_region 38235..38319
/rpt_family="(TTC)n"
repeat_region 38320..38499
/rpt_family="(TTC)n"
```

Qy 169 aaatacgtaaaagcacgaagagtctgtaatacaaacagccaaatatcatgaagagtaccca 228
|| || || || || || || || || || || || || || || || || || || || ||
Db 38579 AAGAAGGAAAAGGAGGAGGAGGAGGAGAGAAGAAGGAAGAAGGAGGAGAGAAGAAGAG 38520

Qy 1 cttctatttggtaaacatggctcataacttttcgtcatccttctcttttccaact 60
|||
Db 28 CTTTCTATTTTGTAACCATGGCTCATAACTTTTGTATCCTTTCTCCTTTTCCAACIT 87

Qy 61 ttactcattactgtctcactaatgatcggtagcacacgctctcgtcagcggtcgacat 120
|||
Db 88 TTACTCATATACGTCTCTACTAATAATCGGTAGTCACACCGCTCTCGTCAGCGGGTCGACAT 147

Qy 121 ttattccacacacaacacctcatcagagtgccacaattgcttcaaaatcagaaaag 180
| | | | |
Db 148 TTATTCGAGACACAACCACTCATCAGAGTGCACAATTGGCTTCAAATACGAAAAG 207

Qy 181 caggaagagtctgaatacaaacacgcacaaatatcatgaagagtaccacaaacatgagaag 240
| | | | |
Db 208 CACAAGAGTCTGAATACAAACACCAAAATATCACGAAAAGTACCCAAAACATGAGAAG 267

Qy 241 cctgaaatgtacaaggagggaaaaacaaaacccctgcaaactcatgaagagtaccacgag 300
| | | | |
Db 268 CCTAAATGCACAAGGAGGAAAAACAAAACCCTGCAAACTCATGAAGAGTACCACGAG 327

Qy 301 tcacgcgaatcgaaaggagcacgaagagtacgataaagaaaaaccggattccccaaatgg 360
| | | | |
Db 328 TCACGCGAATCGAAGGAGCAGGAAGAGTAGATAAAGAAAAACCCGATTCCCCAAATGG 387

Qy 361 gaaaagcctaagagcagcagagaacacgaagtccaatccgaaaaataccggagtacaag 420
| | | | |
Db 388 GAAAAGCCTAAAGAGCACAGAAACACGAAGTTGAATATCCGAAAATACCCGAGTACAG 447

Qy 421 gacaacaaagatg----- 433
| | | | |
Db 448 GACAACAAAGATGAGGTAAGGAACATAAAATGAAGAGTACCATGAATCACGCGAATCG 507

Qy 434 ----- 433
Db 508 AAGGAGCACGAAGAATACGAGAAAGAAAAACCCGAGTCCCCAAATGGGAAAAGCCTAAA 567

Qy 434 ----- 433
Db 568 GAGCAGCAGAAACACGAAGTCAATATCCGAAAATACCCGAGTACAAGGAAAAGCAAGAT 627

Qy 434 -agaataagaacaataaagatgaagagtgccaggagtcacacgaatcgaaagagcacgaa 492
| | | | |
Db 628 AAGAGTAAGGAACATAAGATGAAGAGTGCCACGAGTCACACGAATCGAAAGATCACGAA 687

Qy 493 gagtacgagaaagaaaaaccggattttcccacaaatgggaaaagcctaaagggcacgagaaa 552
| | | | |
Db 688 GAGTACGAGAAAGAAAAACCAATTCTCTCAATGGGAAAAGCCTAAAGAGCACGAGAAA 747

Qy 553 cataaagccgaatatccgaaaaatccctgagtccaaggaaaactagatgaggataaggaa 612
| | | | |
Db 748 CATAAAGCCGAATATCCAAAAATACCCGAGTGCAAGGAAAAACAAGATGAGGATAAGGAA 807

Qy 613 cataaacatgagttcccaagcatgaaaaagaagaggagagaagaacctgagaagggcata 672
| | | | |
Db 808 GATAAACATGAGTTCCTCAAGCATGAAAAAGAGGAGAGAAGAACTGAGAAAGGCAGA 867

Qy 673 gtaccctgagtggttaaatgacctgaatggccgaagtccatgtttactcagtcgtgctc 732
| | | | |
Db 868 GTACCCTGAGTGGGTAAATGCCATGAATGGCCGAAGTCCATGTTACTCAGTCTGGCTC 927

Qy 733 gagcactaagccttaagccatatgacactgggtgatgtgccatcatgcagtaatttc 792
| | | | |
Db 928 GAGCATTAAAGCCTTAAGCCATATGACACTGGTGCATGTGCCATCATCATGCAGTAATTTC 987

Qy 793 atgggatatgtttaatttatattgtaataaaaagatggtagtggtggaatgtgtgtgtgc 852
| | | | |
Db 988 ATGGGATATCGTAATTATATTGTTAATAAAAAGATGGTAGTGGGAAATGTGTGTGTC 1047

Qy 853 attaccatg-agcaatgctgaatctctttgcatgcatagagattctgaatggttatag 911
| | | | |
Db 1048 ATTCAATCATGTAGCAATGCTGAATCTCTTTCATGCATAGAGATTCTGAATGGTTATAG 1107

Qy 912 tttatgttatatggtttgtctagtgaatttaattttgaatgttgtatgtaattgt 967
| | | | |
Db 1108 TTTATGTTATATCGTTTGTCTAGTGAAATTAATTTTGAATGTTGTATCTTAATGTT 1163

RESULT 5
GBU34401
LOCUS GBU34401 1699 bp DNA PLN 01-JAN-1996
DEFINITION Gossypium barbadense FbLate-2 gene, complete cds.

c	25	83.2	8.6	76897	42	AC016179	AC016179 Homo sapi
	26	83	8.6	169931	11	AC005822	AC005822 Homo sapi
c	27	82	8.5	162575	41	AC004086	AC004086 Homo sapi
c	28	81.6	8.4	83440	53	AC024285	AC024285 Homo sapi
c	29	79.6	8.2	80432	51	AC022680	AC022680 Homo sapi
	30	79.6	8.2	222193	32	CNS01DSB	AL121768 Homo sapi
	31	79.4	8.2	3489	82	KSU52064	U52064 Kaposi's sa
c	32	79.4	8.2	32207	5	AR065852	AR065852 Sequence
c	33	79.4	8.2	137508	82	KSU75698	U75698 Kaposi's sa
c	34	79.4	8.2	141753	49	AC009323	AC009323 Arabidops
c	35	79.4	8.2	174383	41	AC009781	AC009781 Homo sapi
	36	78.8	8.1	42839	69	AC027282	AC027282 Homo sapi
	37	78.6	8.1	164520	43	AC020738	AC020738 Homo sapi
	38	78.2	8.1	2000	2	AF019082	AF019082 Borrelia
	39	78.2	8.1	27323	2	AE000789	AE000789 Borrelia
	40	78.2	8.1	110000	31	PFMAL4P1_0	AL034557 Plasmodiu
	41	78.2	8.1	166547	41	HS1164I10	AL049537 Homo sapi
	42	77.8	8.0	43907	34	CELFP36H12	AF078790 Caenorhab
c	43	77.8	8.0	119500	51	AC015927	AC015927 Homo sapi
c	44	77.6	8.0	89072	55	AC025070	AC025070 Homo sapi
c	45	77.2	8.0	73020	51	AC022851	AC022851 Homo sapi

ALIGNMENTS

RESULT 1
I18362
LOCUS I18362 1283 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 17 from patent US 5495070.
ACCESSION I18362
VERSION I18362.1 GI:1598717
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1283)
AUTHORS John,M.
TITLE Genetically engineering cotton plants for altered fiber
JOURNAL Patent: US 5495070-A 17 27-FEB-1996;
FEATURES Location/Qualifiers
source 1. 1283
/organism="unknown"
BASE COUNT 509 a 233 c 251 g 290 t
ORIGIN

Query Match 75.5%; Score 730; DB 5; Length 1283;
Best Local Similarity 82.5%; Pred. No. 4e-140;
Matches 937; Conservative 0; Mismatches 30; Indels 169; Gaps 2;

Qy	1	ctttctatttggtaaacatggtcataacttttgcgtcatcttttctcttttccaactt	60
Db	28	CTTCTATTTTGTAAACATGGCTCATAACTTTGTCACTCTTCTCTTTTCCAACCT	87
Qy	61	ttactcattactgtctcactaatgatcggtgacacacgcgtctcgtcagcggtcgacat	120
Db	88	TTACTCATTACTGTCTCACTAATAATCGGTAGTCACACCGCTCGCTCAGCGGCTCGACAT	147
Qy	121	ttattccac	180
Db	148	TTATTCAGACACAAACACCTCATCAGACGTGCCAATTTGCTTCAAAATACGAAAG	207
Qy	181	cacgaagagtctgaatacaaacagccaaatatcatgaagagtacccaacacatgagag	240
Db	208	CACAAGAGTCTGAATACAAACACCAAAATATCACGAAAGTACCAAAACATGAGAG	267
Qy	241	cctgaaatgtacaaggaggaaaaacacacacacacacacacacacacacacacacacacac	300
Db	268	CCTAAATGCACAGGAGGAAAAACAAACCCCTGCAAAACATCATGAAGAGTACCACGAG	327
Qy	301	tcacgcgaatcgaaggagcacgaagagtacgataaagaaacccgattttcccaaatg	360

Db	328	TCACGCGAATCGAAGGACGACGAAGTAGCATATAAGAAAAACCCGATTTCCTCCAAATGG	387
Qy	361	gaaaagcctaagagcagcagaaacacgaagtcgaatatccgaaatacccgagtacaag	420
Db	388	GAAAAGCCTAAAGAGCACAAGAAACACGAAGTTGAATATCCGAAATACCCGAGTACAAG	447
Qy	421	gacaacaagatg-----	433
Db	448	GACAAACAGATGAGGATAAGGAACATAAAATGAAGAGTACCATGAATCAGCGAATCG	507
Qy	434	-----	433
Db	508	AAGGAGCACGAAGATACGAGAAAGAAAAACCCGAGTTCCTCCAAATGGGAAAGCCTAAA	567
Qy	434	-----	433
Db	568	GAGCAGGAGAAACACGAAGTCGAATATCCGAAATACCCGAGTACAAGGAAAGCAAGAT	627
Qy	434	-agaataagaacataaagatgaagagtgcaggagtcacacgaatcgaagagcacgaa	492
Db	628	AAGAGTAAGGAACATAAAGATGAAGAGTGCCACGAGTCACAGAAATCGAAAGATCAGGAA	687
Qy	493	gagtacgagaaagaaacccgattttcccaaatgggaaagcctaagggcacagagaa	552
Db	688	GAGTACGAGAAAGAAAAACCAATTTCTTCAATGGGAAAGCCTAAAGAGCAGAGAA	747
Qy	553	cataaagccgaatatccgaaatacctgagtcgaagaaacacatagatgaggataaggaa	612
Db	748	CATAAGCCGAATATCCAAATATCCCGAGTCAAGGAAAGCAAGATGAGGATAAGGAA	807
Qy	613	cataacatgagttcccaagcatgaaagagagagagagaaacctgagaaaggcata	672
Db	808	GATAAATCATGAGTTCCCAAGCATGAAAGAGAGGAGAGAAACCTGAGAAAGGCAGA	867
Qy	673	gtaccctgagtggttaaaatgcctgaatggcgaagtcacatgtttactcagtcgtgctc	732
Db	868	GTACCTGAGTGGGTAAATGCCTGAATGGCGAAGTCCATGTTACTCAGTCGCTC	927
Qy	733	gagcactaagccttaagccatatgacactggtgcatggtccatcatcagtaatttc	792
Db	928	GAGCATTAGCCTTAAGCCATATGACACTGGTGCATGTCCATCATCATCAGTAATTTC	987
Qy	793	atgggatattgtaattatattgttaataaaaaagatggtgagtggaatgtgtgtgtc	852
Db	988	ATGGGATATCGTAATTATATTGTTAATAAAAAAGATGAGTGGGAAATGTGTGTGTG	1047
Qy	853	attcatccatg-agcaatgctgaatctcttgcagtcagagattctgaatggttatag	911
Db	1048	ATTTCATCCATGTAGCAATGCTGAATCTTTTCATGATAGAGATTCTGAATGGTTATAG	1107
Qy	912	tttatgttatatcggttcttctagtgaatattttgaatgtgtgtatgtt	967
Db	1108	TTTATGTTATATCGTTGTTCTAGTGAATTAATTTGAATGTTGTATCTAATGTT	1163

RESULT 2
I21349
LOCUS I21349 1283 bp DNA PAT 07-OCT-1996
DEFINITION Sequence 17 from patent US 5521078.
ACCESSION I21349
VERSION I21349.1 GI:1601703
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1283)
AUTHORS John,M.
TITLE Genetically engineering cotton plants for altered fiber
JOURNAL Patent: US 5521078-A 17 28-MAY-1996;
FEATURES Location/Qualifiers
source 1. 1283
/organism="unknown"
BASE COUNT 509 a 233 c 251 g 290 t

```

; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-2

```

Query Match 5.6%; Score 54.2; DB 2; Length 2277;
Best Local Similarity 31.3%; Pred. No. 1.2e-05;
Matches 117; Conservative 66; Mismatches 188; Indels 3; Gaps 1;

Qy 317 agcacaagaagtagcagataaagaaaaaccgcatttccccaaatgggaaaaagcctaaagagc 376
| : | : | : | | | : : : : | : | : | : | : | :
Db 575 ARGAYAAATAYAAAGYACNCARWSNGTNGSNGCNATHGGNGARGARAARWSNCAYARGW 634

Qy 377 acgagaaacacgaagtcgaatatccgaaaataccgcagtagacaaggacaaacaagatgaga 436
| : | | | | : | | : | : | : | : | : | : | : | : | :
Db 635 SNAARCAICARGARWSNGTNAARAARAARGNGMNGARGARGARGAYATGGGARGARGARG 694

Qy 437 ataagaacaataaagatgaagtagtcaggagtagcacgaatcgaaagagcagcaagagtg 496
| | | | : | | | : | : | : | : | : | : | : | : | :
Db 695 ARAAYGAYGAYGAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYGAYGAYGARGATG 754

Qy 497 acgagaagaaaaaccgcatttccccaaatgggaaaaagcctaaagggcagcagaaacata 556
| | : | : | : | : | : | : | : | : | : | : | : | :
Db 755 ARGARGARGARAAAYATHGARW---SNAARGTNACNAARCCNGTNCARATHCARAARMNG 811

Qy 557 aagccgaatatccgaaaatccctgagtgcaaggaaaaactagatgaggataaggaacata 616
| | : | | | : : : : | | | : | : | : | : | : | :
Db 812 CNGTNAARMGNCNGCNCNGCNAARWSNWSNGAYCAYWSNGARGARGAYWSNGAYTTNG 871

Qy 617 aacatgagtgctccaaagcatgaaaaagaagaggagaagaacactgagaaagggcatagatc 676
| : | : : | : | | : | : | : | : | : | : | : | :
Db 872 ARGARWSNGAYWSNATHGAYGAYGGNGARGARYTNGCNCARWSNGAYACNWSNACNGARG 931

Qy 677 cctgagtggggttaa 690
| : | : | : | :
Db 932 ARCARGARGAYAAAR 945

```

RESULT 15
US-09-098-487-2
; Sequence 2, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487

```

Db 19535 AGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTAGAGG 19476
Qy 398 atccgaaaaatcccgagtacaaggacaaacaagatgagaataaagaacataaagatgaag 457
Db 19475 AGCAGGAGGTGGAAGAGCAAGAGCAGGAGGTGGAAGAGCAGGAGCAGGAGCAGGAGAGC 19416
Qy 458 agtgcaggaggtcacacgaatcgaagagcagcaagagtacgagaagaaaaacccgatt 517
Db 19415 AGGAATTAGAGGAGGTGAGGAGCAAGAGCAGGAGCAGGAGGTGGAAGAGCAGGAGT 19356
Qy 518 tccccaaatgggaaaagcctaaggcgacgagaacataaagccgaatatccgaaaatac 577
Db 19355 TAGAGGAGGTGGAAGAGCAGGAGCAGGAGTTAGAGGAGGTGGAAGAGCAGGAGAGC 19296
Qy 578 ctgagtgcaggaaaaactagatgaggataaggaacataaactaggttcccaagatcg 637
Db 19295 AGGAGTTAGAGGAGGTGGAAGAGCAGGAGCAGGAGTTAGAGGAGGTGGAAGAGCAGG 19236
Qy 638 aaaaagaagaggagaagaacacctgagaagcgatcatg 674
Db 19235 AGCAGCAGGGGTGGAACAGCAGGAGCAGGAGACGGT 19199

RESULT 9

US-08-931-999-4

; Sequence 4, Application US/08931999

; Patent No. 6043219

; GENERAL INFORMATION:

; APPLICANT: Iandolo, John J.

; APPLICANT: Crupper, Scott S.

; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hovey, Williams, Timmons & Collins

; STREET: 2405 Grand Boulevard, Suite 400

; CITY: Kansas City

; STATE: Missouri

; COUNTRY: U.S.A.

; ZIP: 64108

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/931,999

; FILING DATE:

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/710,561

; FILING DATE: 19-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Collins, John M.

; REGISTRATION NUMBER: 26,262

; REFERENCE/DOCKET NUMBER: 25043-A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 816/474-9050

; TELEFAX: 816/474-9057

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 6755 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Staphylococcus aureus

; STRAIN: UT0007

US-08-931-999-4

Query Match 6.5%; Score 62.6; DB 5; Length 6755;
Best Local Similarity 50.1%; Pred. No. 1.1e-07;
Matches 211; Conservative 0; Mismatches 204; Indels 6; Gaps 2;

Qy 246 aatgtacaaggaggaaaaaaccctgcaaacatcatgaagagtaccagagtcacg 305
Db 5203 AACGAAAAACGACAGAAACAAACACAAAGAGAGAAAAAGAAAAAACGAGG 6262
Qy 306 cgaatcgaaggagcacgaagagtacgataaagaaaaacccgatttcccaaatgggaaa 365
Db 5263 AGAAACAAA-----AGCAGGGAACAGCAAAACCCGCAAGACGAGCCACAAAAG 6317
Qy 366 gcctaagagcagcagagaacacgaagtgcgaatatccgaaaatacccgagtacaaggacaa 425
Db 6318 AGGAAGAACCGCGAGAAAAAGAAAAAGCCAAAAACAAAGAGAACACAGCA 6377
Qy 426 acaagtatgagataagaacataaagatgaagagtgccaggagtcacacgaatcgaaaga 485
Db 6378 AAAAGAAGAGAGAAAGGGA-AGAAAAAGAGGAGAAACAGAGAGACAAAGCCAAAAAG 6436
Qy 486 gcacgaagagtcagagaagaaaaacccgatttcccaaatgggaaaagcctaaggggca 545
Db 6437 CAACGAAGCGGAGAGAAAGAGCAAAACAAACACGCAAAAGGGCAAAAGACGACCA 6496
Qy 546 cgagaacataaagccgaatatccgaaaatacctgagtgcaggaaactagatgagga 605
Db 6497 AACAAAGACAAACAGCACAACAAAGCGAACACACACCCAAAAAGAGAACAGAA 6556
Qy 606 taaggacataaactagagttcccaagcatgaaaagaagaggagaagaacactgagaa 665
Db 6557 AAACAACACGAGAACAAAGAAAGGGAGGCAAAAGAACAAACAAAAACACGAAGA 6616
Qy 666 a 666
Db 6617 A 6617

RESULT 10

US-08-257-073-4

; Sequence 4, Application US/08257073

; Patent No. 5766597

; GENERAL INFORMATION:

; APPLICANT: Paoletti, Enzo

; APPLICANT: de Taisne, Charles

; APPLICANT: Tine, John A.

; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.

; STREET: 530 Fifth Avenue, 25th Floor

; CITY: New York

; STATE: New York

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/257,073

; FILING DATE: 09-JUN-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/075,783

; FILING DATE: 11-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/852,305

; FILING DATE: 18-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/672,183

; FILING DATE: 20-MAR-1991

```

RESULT      6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Su
; CITY: Alexandria

```

Query Match 10.2%; Score 99; DB 1; Length 7218;
Best Local Similarity 8.2%; Pred. No. 1.7e-17;
Matches 36; Conservative 255; Mismatches 150; Indels 0; Gaps 0;

[illegible]

Db 88 TTACTCATTACTGTCTCACTAATAATCGGTAGTCACACCGTCTCGTCAGCGGCTCGACAT 147
Qy 121 ttattccacacacaaacacotcatcagagctgccacaattggcttcaaaatcagaaaag 180
Db 148 TTATTCAGACACAAACACCTCATCAGAGCTGCCACAATTGGCTTCAAAATACGAAAAG 207
Qy 181 caggaagagtctgaatacaaacacccaaatatcatgaagagtcacccaaatcagagaag 240
Db 208 CACAAAGAGTCTGAATACAAACCAAAATATCAGAAAAGTACCCAAAACATGAGAAG 267
Qy 241 cctgaaatgtacaaggaggaacacacccctgcaaacatcatgaagagtcaccagag 300
Db 268 CCTAAATGCACAGGAGGAAAAACAAACCTCCTCAAAATCATGAAGAGTACCACGAG 327
Qy 301 tcacgcgaatcgaaggagcaggaagagtcagataaagaaaaacccgatttccccaaatgg 360
Db 328 TCACGCGAATCGAAGGAGCACGAAGAGTACGATAAAGAAAACCCGATTTCCTCCCAATGG 387
Qy 361 gaaaagcctaagagcagcagagaacacgaagtcgaatatccgaaaatcccgagtacaga 420
Db 388 GAAAGCCTAAAGAGCACAGAAACACGAAGTTGAATATCCGAAAATACCCGAGTACAAG 447
Qy 421 gacaaacaagatg----- 433
Db 448 GACAAACAGATGAGGATAAGGAACATAAAATGAAGAGTACCATGAATCAGCGAATCG 507
Qy 434 ----- 433
Db 508 AAGGAGCACGAAGAATACGAGAAAGAAAACCCGAGTTCCTCCAAATGGGAAAAGCCTAAA 567
Qy 434 ----- 433
Db 568 GAGCAGGAGAAACACGAAGTGAATATCCGAAAATACCCGAGTACAAGGAAAAGCAAGAT 627
Qy 434 -agaataagaacataaagatgaagagtgccaggagtcacacgaatcgaagagcagcaa 492
Db 628 AAGAGTAAGGAACATAAAGATGAAGAGTGCACGAGTACACGAATCGAAGATCAGGAA 687
Qy 493 gagtacgagaaagaaaaacccgatttccccaaatgggaaaagcctaaggagcagagaaa 552
Db 688 GAGTACGAGAAAGAAAACCAATTTCTCAATGGGAAAAGCCTAAGAGCAGCAGAAA 747
Qy 553 cataaagccgaatatccgaaaatccctgagtgaaggaacatgatgaggataaggaa 612
Db 748 CATAAAGCCGAATATCCAAAATACCCGAGTGAAGGAAAACAAAGTGAAGGATAAGGAA 807
Qy 613 cataaacatgagttcccaagcatgaaaaagagagagagaacacctgagaaaggcata 672
Db 808 GATAAACATGAGTTCCTCAAGCATGAAAAAGAGGAGAGAAACCTGAGAAAGGCAGA 867
Qy 673 gtaccctgagtggttaaatgcctgaatggcgaagtcctgattactcagtcgtgctc 732
Db 868 GTACCTGAGTGGGTTAAATGCCTGAATGGCCGAGTCCATGTTACTCAGTCTGGCTC 927
Qy 733 gagcactaagccttaagccatatgacactggtgcatgtgccatcatcagtaatttc 792
Db 928 GAGCATTAGCCTTAAGCCATATGACACTGGTGCATGTCCATCATCATGAGTAATTC 987
Qy 793 atgggatattgttaattatattgttaataaaaagatggtagtggaatgtgtgtgctc 852
Db 988 ATGGGATATCGTAATTATATTGTTAATAAAAGATGGTGAAGGAAAATGTGTGTGTC 1047
Qy 853 attcaccatg-agcaatgctgaatctcttgcagcatagagattctgaaatgtttatag 911
Db 1048 ATTCATCATGTAGCAATGCTGAATCTCTTTCATGATAGAGATTCTGAAGTTATATAG 1107
Qy 912 tttatgttatatcgtttggttctagtgaataattttgaatgtgtatgtaatttt 967
Db 1108 TTTATGTTATATCGTTTGTCTAGTGAAATTAATTTGAATGTTGTATCTAATGTT 1163

RESULT 4

US-08-298-829-17
; Sequence 17, Application US/08298829
; Patent No. 5620882
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,829
; FILING DATE: 19-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/885,970
; FILING DATE: 18-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1283 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15
; CLONE: E9
US-08-298-829-17

Query Match 75.5%; Score 730; DB 1; Length 1283;
Best Local Similarity 82.5%; Pred. No. 8.9e-188;
Matches 937; Conservative 0; Mismatches 30; Indels 169; Gaps 2;

Qy 1 cttctctatttggttaaccatggctcataactttctcatcctttcttcttccaaactt 60
Db 28 CTTTCTATTTTGTAAACCATGGCTCATAAATTTTGTCTATCTTTCTTCTTCTTCAACTT 87
Qy 61 ttactcattactgtctcactaatgatcggttagccacacgctcgtcagcggtcgacat 120
Db 88 TTACTCATTACTGTCTCACTAATAATCGGTAGTCACACCGTCTCGTCAGCGGCTCGACAT 147
Qy 121 ttattccacacacaaacacotcatcagagctgccacaattggcttcaaaatcagaaaag 180

Qy 733 gagcactaagccttaagccatatgacactgggtgcattgtgccatcatcatgcagtaatttc 792
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Query Match 75.5%; Score 730; DB 1; Length 1283;


```

RESULT 13
AQ026918
LOCUS      AQ026918      890 bp      DNA      GSS      30-JUN-1998
DEFINITION CIT-HSP-2322B22.TF CIT-HSP Homo sapiens genomic clone 2322B22,
            genomic survey sequence.
ACCESSION  AQ026918

VERSION    AQ026918.1  GI:3267140
KEYWORDS   GSS.
SOURCE     human.
            ORGANISM  Homo sapiens
                        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 890)
AUTHORS    Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,

```

Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.

```

FEATURES                                Location/Qualifiers
    source                               1. .890
                                         /organism="Homo sapiens"
                                         /db_xref="taxon:9606"
                                         /clone="2322B22"
                                         /clone_lib="CIT-HSP"
                                         /sex="Male"
                                         /cell_type="Sperm"
                                         /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                                         HindIII"
BASE COUNT      204 a      71 c      44 g      571 t
ORIGIN

```

Query Match 3.5%; Score 107.2; DB 93; Length 890;
Best Local Similarity 49.0%; Pred. No. 6.6e-07;
Matches 344; Conservative 0; Mismatches 353; Indels 5; Gaps 2;

Qy 196 tcaatgaatcgatttccaatttttcgcagtataagttccttttaactctttctttttacttc 2023
| | | | | | | | | | | | | | | | | |
Db 147 TTATTAATTTTATTTTTTTTTTTATTATTATAATTTTTTTTATTTTTTTATTAAATTA 206-
| | | | | | | | | | | | | | | | | |

Qy 2024 attttataacgaattctatgggataatgtccctacaacatgcatcatacagtgtttaatt 2083
| | | | | | | | | | | | | | | | | |
Db 207 TTTTTTTTATATTTTT---ATTATTTTATTTTTATTTTAAATTAATTTTTTTATT 262
| | | | | | | | | | | | | | | | | |

Qy 2084 ataaattccattctctatattttactaagatataggtaacctcaaacgtgctgatttttact 2143
| | | | | | | | | | | | | | | | | |
Db 263 ATATTTATTTTTTTTTTATTTTATTATTATTATTATTATTATTATTATTATTATTATTAT 322
| | | | | | | | | | | | | | | | | |

Qy 2144 aattttattattataaaattgtagaatgattatttttcaataaatttaacacaatattta 2203
| | | | | | | | | | | | | | | | | |
Db 323 TTATTTATTTTATATTTTTTTA-TATTTTTTTTTTTAAATTTTATATTATTATTATTTT 381
| | | | | | | | | | | | | | | | | |

Qy 2204 atattattattattattattttctcaatttttattaacaaaaacataaatttttgacaaa 2263
| | | | | | | | | | | | | | | | | |
Db 382 ATATATTATTTTTTTTTTATTTTTTATTTTATTATTATTATTATTATTATTATTATTTT 441
| | | | | | | | | | | | | | | | | |

Qy 2264 ttaaanaataatgaattaaatttctcaatttttcgtgcaactattacaaaaatccttcatag 2323
| | | | | | | | | | | | | | | | | |
Db 442 TTTTTTTTATTATTTTTTTTTTATTATTATTATTATTATTATTATATATAATATATTTTATAT 501
| | | | | | | | | | | | | | | | | |

Qy 2324 tcctaattctaaatttgatgcagagggtgataataactctaaatttgatgcagaggtaataat 2383
| | | | | | | | | | | | | | | | | |
Db 502 TATTTAATATATTTTTTATTTTATTTTATATTAATTATTATATATTTTAATTTATTA 561
| | | | | | | | | | | | | | | | | |

Qy 2384 gggccgggttgagctggacttaagcatgatattgacgtacttttatattttccaaattc 2443
| | | | | | | | | | | | | | | | | |
Db 562 TTTTTTATTTTAATTATTTTTTTTTTATTATTATTATTATTATTATTATTATTATTATT 621
| | | | | | | | | | | | | | | | | |

Qy 2444 aaccagctcgaaatatgagtcctaaaattttgccaatthaatccaagccccattttaagt 2503
| | | | | | | | | | | | | | | | | |
Db 622 TTTTTTAATTTTATTATTATAATATAATTTTTTATATAATTCATTATTAATTTTATA 681
| | | | | | | | | | | | | | | | | |

Qy 2504 tcgtccatattatttttaatttaaaaaatttatatcatcttttattttaattttaattat 2563
| | | | | | | | | | | | | | | | | |

Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Malvales;
Malvaceae; Gossypium.

REFERENCE 1 (bases 1 to 921)

AUTHORS Leslie, A., Frisch, D., Yu, Y., Wood, T. C., Wing, R. A. and Wilkins, T. A.

TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
High quality sequence stop: 921.

FEATURES

source Location/Qualifiers

1. 921

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db_xref="taxon:29729"

/clone="GA_Ea0023023"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"

/lab_host="E. coli"

/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 265 a 185 c 190 g 281 t

ORIGIN

Query Match 3.7%; Score 113.8; DB 80; Length 921;
Best Local Similarity 67.6%; Pred. No. 7.2e-08;
Matches 209; Conservative 0; Mismatches 87; Indels 13; Gaps 3;

```
Qy    952 ctacttaataatagataaaataattgtggtacattagatcaaagaacaaactagatttt 1011  
      || | | | | | | | | | | | | | | | | | | | |  
Db    715 CTTTATTAATAATTGAGAAAATTAGTCCTGTACGTTAGATCAAAACAGCAAATTAACATT 656  
  
Qy   1012 gtcccattctattgtttaaaagtcggtcggttacattaaaataagggtacatgtttacatgc 1071  
      || | | | | | | | | | | | | | | | | | | | |  
Db    655 -----TTATTAAAAAATGGTCCTGTACATCAACATAAGGTACACATGGCATGC 607  
  
Qy   1072 caccgtataactatctggttattctcatcaatcacgctaatttttaacagtagamaatgaatg 1131  
      || | | | | | | | | | | | | | | | | | | | |  
Db    606 TACGAGTCACTACTAGTTATTCTGTGCAACCATAACCGGTTTTAACCAATATAAATGGATG 547  
  
Qy   1132 taattttttaaatagaagggtc aaattgttatttgatctaacacgtaggattaatttac 1191  
      || | | | | | | | | | | | | | | | | | | | |  
Db    546 GAATTTTTTAACAAG-AAGAACCAATTTGCTCTTTAATCTAATATGCAGGGACTAATTTGTT 488  
  
Qy   1192 ttttttccctaagaanaataagtaaaatatatttgaatcttaatacaaaaaactttcatga 1251  
      || | | | | | | | | | | | | | | | | | | | |  
Db    487 CTATTTT-ATGAGTATAGGGGCCAAATGCAATCTGACTCCTAGTACAAGACTTCCATAG 429  
  
Qy   1252 tacttttat 1260  
      || | | | | |  
Db    428 TACCTTTAT 420
```

```

RESULT 10
CNS003BD/c
LOCUS      CNS003BD      1101 bp      DNA              GSS              03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL064091
VERSION    AL064091.1  GI:4941847
KEYWORDS   GSS.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```

```

Muscomorphia; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
                - Web : www.genoscope.cns.fr)
COMMENT        Determination of this BAC-end sequence was carried out as part of a
                collaboration with the Berkeley Drosophila Genome Project (BDGP).
                The BDGP is constructing a physical map of the Drosophila
                melanogaster genome using these BACs. For further information
                please see http://www.fruitfly.org The BDGP Drosophila
                melanogaster BAC library was prepared by Kazutoyo Osogawa and
                Aaron Mammossier in Pieter de Jong's laboratory in the Department of
                Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                NY. The library is named RPCI-98 and was constructed by partial
                EcoRI digestion of Drosophila DNA provided by the BDGP from the
                isogenic strain y2; cn bw sp, the same strain used for the BDGP's
                P1 and EST libraries. A more detailed description of the library
                and how to order individual BAC clones, the entire library, or
                filters for hybridization from the BACPAC Resource Center can be
                found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
    source      Location/Qualifiers
                1..1101
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone_lib="RPCI-98"
                /clone="BACR08K08"
                /note="end : TET3"
BASE COUNT     395 a    120 c    103 g    334 t    149 others
ORIGIN

```

Query Match 3.6%; Score 109.6; DB 122; Length 1101;
Best Local Similarity 40.5%; Pred. No. 2.9e-07;
Matches 247; Conservative 73; Mismatches 289; Indels 1; Gaps 1;

[illegible]

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES             Location/Qualifiers
     source           1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_lib="RPCI-98"
                        /clone="BACR08K08"
                        /note="end : TET3"
BASE COUNT           395 a    120 c    103 g    334 t    149 others
ORIGIN

```

Query Match 3.9%; Score 118.4; DB 122; Length 1101;
Best Local Similarity 42.2%; Pred. No. 1.5e-08;
Matches 234; Conservative 61; Mismatches 260; Indels 0; Gaps 0;

Qy 383 aacctttgaaagtcaccaaaccttttaacaaagtttgattgtgtacatatatatatatct 442
| ||| ||| : | ||: : || ||| || : : || : |
Db 510 ACCTTTAATAAAWAAAAAAWAAAAAAWTTTTTTTTWWTMTTTATAMATAAAAA 569

Qy 443 tcaaatatttataataaaattgtgtttaaataatttacagttatattattttttatctc 502
||| | || |||| | || |: | | | | |||| | |
Db 570 AAAAAAAAAAAAAAAAAAAAAAAAAAAWWTMTTTTTTTTTTTTTTTTTTTTTT 620

Qy 503 taattttatttgcgcgaattttttagttgatattttaacataaaaaaattgtacacat 562
: ||||: | | ||||| | |: ||| |: |||: | | |
Db 630 AWTTTTTTWTATNTTWTTAATTTTAAATAAWTTTATTATWNAAWATTAAAAAATA 689

Qy 563 ttacaagcccatatacaaaataattataataaattattcattaaaaatatatttaaatatag 622
| | | |||| || | | ||||| : | | | | |: | : |||||
Db 690 AAAAAAAAAATATAAAAAAAAAATATAAAWTTWAAATATATAWTTTAWAAATATAT 749

Qy 623 gatataaataaactatttttagaattattctactttaagataacataggttaaatgtata 682
| | |||: ||: | | | ||| |: : | | || : : | | :
Db 750 TAAAAAAWATAWTTTATTATATTAATAATATWATWTTTATATATNWNWATAWTTW 809

Qy 683 attaataaggttagtttattgtgaagatgagtatatgtcgtaacataatcactaac 742
||| | | ||| |: | : | : | : | | |
Db 810 TTTAAATTTWATWTTTAAATWMTTAAAWATTNTNAWAATAAAAAWAAAAAATAATAA 869

Qy 743 atttttataactctcttggttttgaagttccaaaaagaaaatggaagggaatttgagag 802
| | : | : : | | : | : | : | | | | | : | |
Db 870 AATAATWTATWATWTTTAAATTAATWTAATWTAWAAAAAATAATWATAAATA 929

Qy 803 taagttcatgtttatattatacataatgaagttgatgttttctcttttaattttta 862
: || | | | | | : | : | : | | | | | : | |
Db 930 WAAWTAWTTWTTTMTTAWAAAWWTATWAWAWAATATTTTATATWATWATATAATA 989

Qy 863 tacaaaatttttaataaaaataaataaggttgatgaataataatgaagtcggttt 922
: || | | | | | : | : | : | : | : | | : :
Db 990 WAAAAAATAAATAAATWAAATNNAWAWWAYAAAWAYACWAAAAWTMTATWYATAWA 1049

Qy 923 actaatagtcataatt 937
| | | | | | |
Db 1050 AAAAAAATAATANATT 1064

```

RESULT 7
CNS003DQ/c
LOCUS   CNS003DQ      1101 bp      DNA           GSS           03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08109 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.

```

```

ACCESSION      AL064580
VERSION        AL064580.1  GI:4941932
KEYWORDS       GSS.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
                Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 1101)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
                BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                - Web : www.genoscope.cns.fr)
COMMENT        Determination of this BAC-end sequence was carried out as part of a
                collaboration with the Berkeley Drosophila Genome Project (BDGP).
                The BDGP is constructing a physical map of the Drosophila
                melanogaster genome using these BACs. For further information
                please see http://www.fruitfly.org The BDGP Drosophila
                melanogaster BAC library was prepared by Kazutoyo Osoegawa and
                Aaron Mammoser in Pieter de Jong's laboratory in the Department of
                Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
                NY. The library is named RPCI-98 and was constructed by partial
                EcoRI digestion of Drosophila DNA provided by the BDGP from the
                isogenic strain y2; cn bw sp, the same strain used for the BDGP's
                P1 and EST libraries. A more detailed description of the library
                and how to order individual BAC clones, the entire library, or
                filters for hybridization from the BACPAC Resource Center can be
                found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

```

```

FEATURES             Location/Qualifiers
     source            1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone_lib="RPCI-98"
                        /clone="BACR08I09"
                        /note="end : TET3"
BASE COUNT           291 a    51 c   117 g   404 t   238 others
ORIGIN

```

Query Match 3.8%; Score 115.8; DB 122; Length 1101;
Best Local Similarity 44.2%; Pred. No. 3.6e-08;
Matches 253; Conservative 45; Mismatches 272; Indels 2; Gaps 1;

Qy 402 accttaacaaagttgttggtgatcatatatatatcttcaaaattttataataaaa 461
| | | | | : : : | : | | | | : : : | | : | | |
Db 574 ATTTTACTTTTTTYYCYTCNWAYAAWAAATATWAAATMACWTWAAATTTTWTATTAAT 515

Qy 462 ttgtgtttaaataatttacagttatattattttttatctctaattttattgtgcgcaa 521
| | | | | | | | | | | : | : | | | | | | : :
Db 514 TTAATATTTTAAATTTTTTATTTCWTATTTTTTCCCWTTTTTATATTTTAATWW 455

Qy 522 attttttagtgatatttttaacataaaaaaaattgtcacacatttacaagcccatatacaaa 581
| | | : | : | | | : | : | | | : : | : | | | | |
Db 454 AATTAWATTWAAAATTAWAWWAAWTATAAWAAWTAAAAAWAAATAAAATATATAAAAA 395

Qy 582 taattatataaataattcattaaaaaataattttaaataaggatataaataaactattt 641
| | | | | | | | | | : : : | | | | : | | | | | | : |
Db 394 TATTTTAAAAATA--AATATTTWTTTTTTATATAAAWAAATATATAAATATAWAAWAT 337

Qy 642 tagaattattctacttttaagataacataggttaaatgtataaataaagggttagtttat 701
| : : | | | | | : : | : | | | | : | : : : |
Db 336 TWWAWTATATTTTTTATATATTTTWTATTTTTTTWTATTTANTTTTWTWTATAWAWT 277

Qy 702 tgtaagatgagtgatatatgtcgtaaacataatcactaaccattttttataaactcttgg 761
: | | | | | : | | | | | | | | | | |
Db 276 WATTAAATTAATTACAawaaaaaaataaaaaaaataaaaaataaaaatttAAAA 217

Qy 762 ttttgaagttccaaaagaaaaatggaagggaaatttgagagtaagttcatgtttatatta 821
| | | | | | | | | | | | | | | | | | | | |
Db 216 TTTAAAAATAAAAAAaaaaaaacataaaaaaaataaaaaaaataaaaaaaataaaaa 157

Qy 822 tacataatgaagttgatgttttctcttttaataatttttatacaaaattttaataaaa 881

Qy 312 cacatatcacattttgagtttgatgatgataagtcgcataaancgaaatattggtgtgat 371
|::: :|::|:| :|: : : :| |::| :| :|:
Db 442 CMMMMMMHAMATYYTCTCAHTWTMMMMMMWAAATWTWAAAAWAAWTTATWAATWAAAAAW 501
Qy 372 ottcacttttgaactttgtataagtcacccaacttttaacaaagtgttattgtgtacatata 431
:: : ||| : : :| | : | : : : ||| || : | :|:
Db 502 AWWWATTATTTTWWWWTTWATTTTWTWAWTWTATWAAAAAAAWATAATTTAAAWWAAT 561
Qy 432 tatatatattctcaaattttataataaaaaattgtgtttaataatttacagttatattat 491
: | : | | | : ||| || : | | :| | ||| | | ||| |
Db 562 AWATATAAWAATTTAWAAWWTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTT 621
Qy 492 ttttttatctctaattttttattgtgcgcaaaattttttagttgatattttaacataaaaaaa 551
||: | | | | ||| ||| : : :|::: | ||| ||
Db 622 TTWATAAAATTTTAAATAATTTAAATTATWTATTAATAAWTWATTTWWTAAATTAATAAT 681
Qy 552 attgtacacatttacaagcccatatacaaaataattataaataattcattaaaaaatata 611
|: | | | | | | :| :|::|::|::|: | :| | :| | | :|
Db 682 TTWAAATAWAAAAAAAAAAAAAAWATAWAAWATWATAWATAAATTTAAAWATAAAAAWA 741
Qy 612 tttaaatataggatataaaatataaactattttagaattatttctacttttaagataacatagg 671
: | : :| | |||::| ||| : |::| | : : : : :| :| | :| :|
Db 742 AWAATWAWWATAATATWATAT-ATAWTTWAWWATWAAWWTATATAWATAWATAWAAW 800
Qy 672 ttaaagtataaattaataagggttagtttattgttaaagatgagtatatatgtcgtaaacat 731
:|::| ||| |::: : : : : :| : : |::|: : :| :|
Db 801 AWAATAAATAWATAWATAWAAWAAWAWATAWAAWATATAWAAATAWAAAAAAATWAA 860
Qy 732 aatcactaacaccatttttataaacttcttggttttgaagttccaaaaagaaatggaagg 791
|| | | : | |::: : :| |::|: | | : :| : :| :|
Db 861 TATWAATWATAWAAAAAAATAWATWTWTTTWWAAWATATAAAWAWATAWAAWAAAA 920
Qy 792 aaatttgagagtaagttcatgtttatattatcacataatgaagttgatgttttcttctttt 851
|| | | | ||| : :|::| | | |::|: | | : :| : :| :| :|
Db 921 AAAAAAAAAATAWAAWWTWATATTTWTATTAATWTATWATWATWATWATWATWATW 980
Qy 852 taatattttttatcaa-aatattttaataaaaaataaaggattgaatgaaaaatataat 910
|:|::| | ||| : :|::| | | |::|: | | : :| : :| :| :|
Db 981 TTWATWATATATATWTAWATAWATATATTTATTAATWTATATTTTAAWAAATATATATW 1041
Qy 911 gaaagtcgttttactaatagtcataattgcatcttctacttacttaataatagataa 970
| : : : | :|::| | | :| | :| :| :| :| :| :| :| :|
Db 1041 ATAWWWATWATATAWAAWAAATTAATWTATATATTAATAWATAWAAAAAAWATAWAAW 1101
Qy 971 a 971
|
Db 1101 A 1101

```

RESULT 4
BI2681
LOCUS      B12681      804 bp      DNA      GSS      14-MAY-1997
DEFINITION F27D1-Sp6.1 IGF Arabidopsis thaliana genomic clone F27D1,
genomic survey sequence.
ACCESSION  B12681

VERSION    B12681.1 GI:2093801
KEYWORDS   GSS.
SOURCE     thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
            Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 804)
AUTHORS    Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
TITLE       BAC End Sequences at ATGC
JOURNAL     Unpublished (1997)
COMMENT     On Dec 15, 1999 this sequence version replaced gi:4123046.
            Other_GSSs: F27D1-Sp6.2, F27D1-Sp6
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends
            High quality sequence start: 335
            High quality sequence stop: 346.

FEATURES             Location/Qualifiers
     source            1..804
                       /organism="Arabidopsis thaliana"
                       /strain="Columbia"
                       /db_xref="taxon:3702"
                       /clone="F27D1"
                       /clone_lib="IGF"
                       /sex="hermaphrodite"
                       /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                       Produced by Thomas Altmann"

BASE COUNT      241 a      14 c      20 g      439 t      90 others
ORIGIN

```

Query Match 3.9%; Score 120; DB 120; Length 804;
Best Local Similarity 49.8%; Pred. No. 9.3e-09;
Matches 314; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

Qy 2046 taatgttcctacaaacatgctcattacaagtgtttaataaaaattccattotcttatttt 2105
| | | | | | | | | | | | | | | | |
Db 172 TTTNNNNNTANAAAAATTTNNTTTTTTTTTTTTTTNNNAATAAATTNANTTTAAAAATTAA 231

Qy 2106 actaagatattagtaacttcaaactgctgatgttttactaatattattattataaattggt 2165
| | | | | | | | | | | | | | | | |
Db 232 AATATAATTTTAATTTTTTAAAAAATTTTTTATTATTTTTTNATTNTTTTTTNATTTT 291

Qy 2166 agaatgatattttttcaataaatttaacaacaatatttaaattattattattattatttc 2225
| | | | | | | | | | | | | | | | |
Db 292 ---TNNTTTTTATTAATAATTTTAAATNTATATTA AAAAATATTTTTTAAAAATTTT 347

Qy 2226 tcaattttttattaacaaaaacataaatttttgacaaattaaataaataaataaatttc 2285
| | | | | | | | | | | | | | | | |
Db 348 TAATTNAAAATAAAAAAAAAAATTTTATTTTATTATAATAATTA AAAAANTNTATATTTAT 407

Qy 2286 tcaatttttcgtgcaactattacaaaaatccctcatagtcctaactttaatttgatgcag 2345
| | | | | | | | | | | | | | | | |
Db 408 TTTTTTTTTTTTTTAAANNNTATTAANAATAATTANAATTTTTTNAATTTTNNAAATAA 467

Qy 2346 aggtgataaatactttaatttgatgcagaggtaataatgggccgggttgagctggacct 2405
| | | | | | | | | | | | | | | | |

.....

```
Qy      370 atcttcacttttgaaccttgataagtcaccaaacctttaacaaagtgtgattgttacata 429
        ||| | | | | | | | | | | | | | | | | | | | | | | |
Db      648 ATTTTTTCCTTCTCACTTGGCAAATACAATTCTCGAGATCAATAACCTCGTCTTTTAAT 589

Qy      430 tatatatatatcctcaaattttatataaaaaattgtgtttaataatttacagtatatatt 489
        || | | | | | | | | | | | | | | | | | | | | | |
Db      588 TTTTTCCTCGTCTTTTAACTATTATAAAAATATTGAATTATAAAAATATGTAATTATA-A 530

Qy      490 atttttttatctctaattttatttgtgcgcaaaatttttagttgatattttaacataaaaa 549
        || | | | | | | | | | | | | | | | | | | | | | |
Db      529 ATACTTTAATTATAAAAATATGTAATTATAAACTTTTAATTATAAAAATATGTAATTATAA 470

Qy      550 aaattgtcacatttaacaaggcccatatacaaaaattatataaatattcattaaaaaata 609
        | | | | | | | | | | | | | | | | | | | | | | |
Db      469 ATACTTTATAAAAATATGTAATTATAAAATATGTAATTATAAACATTTTAATTATAAAAAT 410

Qy      610 tatttaaataataggatataataataactattttagaattattctacttttaagataacata 669
        || | | | | | | | | | | | | | | | | | | | | | |
Db      409 TGTAATTATAAACATTTTAATTATAAAAATATGTAATTATAAACATTTTAATTATAAAAAT 350

Qy      670 ggttaaattgtataattaataaggttagtittattgttaaagatgagtatatatgtcgtaaac 729
        || | | | | | | | | | | | | | | | | | | | | | |
Db      349 TG-TAATTATAAACATTTTAATTATAAAAATATGTAATTATAAACATTTTAATTATAAAAAT 291

Qy      730 ataatcactaaccattttttataacttcttggtttgaagttccaaaaagaaaattggaag 789
        || | | | | | | | | | | | | | | | | | | | | | |
Db      290 ATTAAATTATAAACATTTTAATTATAAAAATATTTAATTATAAAATATTTAATTATAAAAAT 231

Qy      790 ggaatttgagagtaagttcatgtttatattatatacataatgaagttgatgtttcttctt 849
        || | | | | | | | | | | | | | | | | | | | | | |
Db      230 ATTTAATTATAAAATATTTTAATTATAAAAATATTTAATTATAAAATATTTAATTATAAAAAT 171

Qy      850 ttttaattttttatatacaaaatatttaataaaataaataaggattgaatgaaaaatataa 909
        || | | | | | | | | | | | | | | | | | | | | | |
Db      170 ATTTAATTATAAAATATTTTAATTATAAAAATATTTAATTATAAAATATTTAATTATAAAAAT 111

Qy      910 tgaaagtcgttttactaatagtcatattgcattttgtcgcatctacttaataatagata 969
        || | | | | | | | | | | | | | | | | | | | | | |
Db      110 ATTTAATTATAAAATATTTTAATTATAAAAATATTTAATTATAAAATATTTAATTATAAAAAT 51

Qy      970 aattaaattggtgtacattagatcaaagaa 998
        || | | | | | | | | | | | | | | |
```

Db 50 TTTTAATTATAAAATATTTAATTATAAAA 22

RESULT 14

```

US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
US-07-867-106-2

```

Query Match 2.4%; Score 74.4; DB 1; Length 5852;
Best Local Similarity 48.3%; Pred. No. 4.9e-05;
Matches 277; Conservative 0; Mismatches 286; Indels 11; Gaps 2;

Qy 406 taacaaggttgattgtgacatatatatatatcttcaaatTTTataataaaaaattg 465
||| |||| | || | || |||| | || | || |||| | || |
Db 5821 TAATAAAGACTCTTTAAATTTATTAATATACATTTTAAATGGTATTAAATTTATTTAAAT 5762

Qy 466 gtttaataatttacagttatatttttttatctctaattttatttgcgcaaatTT 525
|| | || || | || | || | || | || | || | || | || |
Db 5761 TATTTGTATTGTATTTTTTATATATGTATTGTGTGTTGTTTTTACTTATATT 5702

Qy 526 ttagtgtatTTTtaacataaaaaaaattgtacacatttacaagcccatatacaataat 585
|| | |||| | || | || | || | || | || | || | || |
Db 5701 CTATTTTTTATTATATAAAATTAATTAATTAATTTTAAATAAATAAAAAAAAAAAAAAAAAA 5642

Qy	2183	ataattttaacaacaatatatttaattattattattattttctcaattttttattaaca	2242
Db	1505	ATATTCAATTATATTGTGATAAAATTTTTTAAATAATTATAAATTTCTTTATATGATTGATA	1446
Qy	2243	aaaacataaaattttgcacaaataaaaaaatgaattaattttctcaatttttcgtgcaac	2302
Db	1445	ATCACATAAATGAGTTATACTTTTTTGGAAATTTTATCAAATGTATATTATTTTTTTTAAC	1386
Qy	2303	tattacaaaaatccttcatagtcctaattcttaatttgatgcagagggtataataatctta	2362
Db	1385	ATTGAAAGATATATCTAAATTTTTTAAATTAAATAATTTTCTATAAAATTTATAATCTTT	1326
Qy	2363	atttgatgcagaggtaataatgggcgggtttgagctggacttaagcatgatattgacgt	2422
Db	1325	TATTTGTATTGATTCCATATTTAACCTCAACTATACTAATAGGAAAAACATTATTAAGTT	1266
Qy	2423	actttatatttttccaattcaaccagctcgaaatatgagtcataaaattttgcgaatt	2482
Db	1265	ACCAAAATTTATTTTGTAGATATTATTAATTTTTTAAATTTACTAAATTTATAATAAAAT	1206
Qy	2483	taatccaagccattttaagttcgtccatattatttttaattaaaaaatttatcat	2542
Db	1205	ATAGTCATTATATATACATGATTACAAATTTAAAAATTTCTATAGAATGTGGTAGTATAAT	1146


```

RESULT 6
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline

```


Query Match 3.2%; Score 98.8; DB 4; Length 19124;
Best Local Similarity 45.3%; Pred. No. 3.7e-09;
Matches 561; Conservative 1; Mismatches 659; Indels 18; Gaps 5;

Qy 139 atataataaacatcgtgagaantaatttttccaattgaagctttaaccactcttaa 198
| | | | | | | | | | | | | | | | | |
Db 7210 ATATATGTATATTACAGTAGTATTATAATGGTAGAATAAGAATAAACACTTTTG 7151

Qy 199 tatttgtagatgtaatttaaatgaagaataaaacatatcttctggacagtattttcatc 258
| | | | | | | | | | | | | | | | | |
Db 7150 AATGTATATATATATGTAAGGTATAATTTATGTATTACAATATATAAAATATT---GTA 7094

Qy 259 ttaattgtttggttcgtgatagggtgatgtcacgtctctttaaatacacatat 318
| | | | | | | | | | | | | | | | | |
Db 7093 TATATATATATATATATTAATAGTTGTACTATTATAATATTACAATATATGTATATGT 7034

Qy 319 cacatttgtggttgatgatgaagtcgcacataancgaataatggtggtgatottcact 378
| | | | | | | | | | | | | | | | | |
Db 7033 TAAAAAATAATATTTAATATGTATATAATAATAATATTATTAGTTTTATATATTTTTAAA 6974

Qy 379 ttggaactttgataagtcaccaaaacttttaacaaagtgttattgtgtacatatatatat 438
| | | | | | | | | | | | | | | | | |
Db 6973 AAAATATATATATATATTAATAAATTTATAATAAATTTAAATATTCTAACAAAAAAAAAA 6914

Qy 439 attctcaaattttataaaaaattgtgtttaaataatttacagttatatattttttta 498
| | | | | | | | | | | | | | | | | |
Db 6913 TATAATCAGAAATATTATATTTTATGTATTCCCTTTATTTCATTTAATTATATATATA 6854

Qy 499 tctctaatttttttgtcgccaaattttttagttgatatttaacataaaaaaaaaattgtac 558
| | | | | | | | | | | | | | | | | |
Db 6853 TTATTTTTTTTATGTTT-----TATTTATTAAGTAAAAATTATAATGAGAAAAAAAAAAT 6799

Qy 559 acatttacaagcccatatacaaaaattataataattcattaaaaaataattttaaat 618
| | | | | | | | | | | | | | | | | |
Db 6798 ACGAAAAACAAACATATAAAAAAGTATATATGCACCGTGTATATATTTAATTATTAA 6739

Qy 619 ataggatataaaataaactattttagaattattctactttaagata---acataggttaa 675
| | | | | | | | | | | | | | | | | |
Db 6738 CATTAATATATGTATATTTTTTTTGACTTTATTTTAATTTATATATATATATATATT 6679

Qy 676 atgtataaatttaaaaggttagttttattgttaagatgagtatatatgtcgttaacataatc 735
| | | | | | | | | | | | | | | | | |
Db 6678 AGAGATAACAAAAAGAGACAAATATGTTCTCGTTTCTCTATCTATATTATCTATCATATA 6619

Qy 736 actaaccatttttattaactcttctgtgtttgaagttccaaaaaagaattggaaggaaat 795
| | | | | | | | | | | | | | | | | |
Db 6618 TTTATATATTATATATATATAATTGATATAGATACATATTCTTTGTATTGTTGATTAT 6559

Qy 796 ttgagagtaagttcatgttttatattacataatgaagttgatgttttctctcttttaaat 855
| | | | | | | | | | | | | | | | | |
Db 6558 ATTAAAGTAGTATATTATTATTAAATTTTGTGTTATATATTATAATTATTATATAT 6499

Qy 856 atttttatacaaaatatttaataaaa---ataattgaagattgaatgaaaaataaatg 911
| | | | | | | | | | | | | | | | | |
Db 6498 AATATATATATAGCATCAAAAAAAAAAATGATAATAATAACAGGAAAAATATATTATT 6439

Qy 912 aaagtcgttttactaatagtcattgttcattttgtcgcatctacttaaatagatataaa 971
| | | | | | | | | | | | | | | | | |
Db 6438 ATATTATATTATATATATTAATAAAAAATGTTTTATCATTGTTTGTGTTGTTATTTTT 6379

Qy 972 ttaattgtggtacattgatgacaagaacaaactagattttgtcccattctattgtttaaaa 1031
| | | | | | | | | | | | | | | | | |
Db 6378 ATGTATTTCATGCATTTTATGAATTTCAAATTTTATTGTATAATATAAAAAATAAGTA 6319

Qy 1032 gctggtccgtttacattaaaaataaggtacatggttacatgccacgtataactatctggtta 1091
| | | | | | | | | | | | | | | | | |
Db 6318 AAAATACACATTATATAATATATATATTCAAATATGAGTTTATTAATAAATGTTC---TG 6262

Qy 1092 ttctatcaatcacgctaatttttaacagtagaaatgaattttttaaatagaagg 1151
| | | | | | | | | | | | | | | | | |
Db 6261 TTCTATATATTATATAAATGAAAAATTTGTTATAATAATAATACATATATGCTACTAT 6202

Qy 1152 tcaaattgtttattgatctaacacgtagggatttaattacttatttttctaaagaataa 1211

```

      ||| | | | ||| | | | | | | | | | | | |
Db  6201 ATAAATATTATAATATCTTTAAAGTATATACTAAAATATATAAAAAATGCATGTATAAA 6142
      ||| | | | | | | | | | | | | | | | |
Qy  1212 gtaaaaataaattgaaacttaatacaaaaactttcatgacttttatcatattttact 1271
      || | | | | | | | | | | | | | | | |
Db  6141 ATAGTATAAAATCATACATATATATATATATATATATATATATATATATATATATA 6082
      ||| | | | | | | | | | | | | | | | |
Qy  1272 tataatttaaatattgtgagagtaacaaraatataaaacatagaacaccaaagttagtt 1331
      ||| | | | | | | | | | | | | | | | |
Db  6081 TATATATGCATATATGTAGAATAAAATTTTATATTTCCAATACTGATATTGTTTATA 6022
      ||| | | | | | | | | | | | | | | | |
Qy  1332 atgggtgtgactcatatacacagtgtaaaattgaataaat 1370
      ||| | | | | | | | | | | | | | | | |
Db  6021 TTTGTATATTTATAATAACAAAAAAGAACGACAAGAAGT 5983

```

RESULT 2

```

US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
;   APPLICANT: Sim, Kim L.
;   APPLICANT: Chitnis, Chetan
;   APPLICANT: Miller, Louis H.
;   APPLICANT: Peterson, David S.
;   APPLICANT: Su, Xin-zhaun
;   APPLICANT: Wellems, Thomas E.
;   TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
;   TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;   NUMBER OF SEQUENCES: 45
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Knobbe Martens Olson & Bear
;     STREET: 620 Newport Center Drive 16th Floor
;     CITY: Newport Beach
;     STATE: California
;     COUNTRY: US
;     ZIP: 92660
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/487,826B
;     FILING DATE: 10-SEP-1993
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Israelsen, Ned
;     REGISTRATION NUMBER: 29,655
;     REFERENCE/DOCKET NUMBER: NIH121.001CP1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 235-8550
;     TELEFAX: (619) 235-0176
;   INFORMATION FOR SEQ ID NO: 13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 19124 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO
US-08-487-826B-13

```

Query Match 3.2%; Score 96.2; DB 4; Length 19124;
Best Local Similarity 45.1%; Pred. No. 1e-08;
Matches 591; Conservative 0; Mismatches 709; Indels 10; Gaps 6;

Qy 138 aaataataaaatcacgtagaaaataaaattttatccaattgaagtct---taaccatct 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5418 AATACGTAACCTGTATTATAGAAAATAAAGAATTAAATATTAAGGAATAAATAAATAT 5477

Db 538 ATATTAGCTATAGTCTTGCTTGCCTAGTTCTAGAAAATTATATATATAAATTATAA 479
Qy 1141 aa-----tagaaagggtcaaaattgttattgatctaacacgtagggaattacttat 1195
Db 478 TACTTTTTTAATAAGCTATTAGTAAAAATTAATATATTATCTAAGTATATAAGGCAAT 419
Qy 1196 ttctctaaagaataagtaaaatataattgaattcttaatacaaaaacttt 1246
Db 418 TTATTAGAAAGTTATGTAGATTAGTTAATATAATATTAGAAATATACTT 368

RESULT 14
Q11710

ID Q11710 standard; DNA; 5852 BP.
AC Q11710;
DT 30-JUL-1991 (first entry)
DE Dictyostelium plasmid Ddp2 containing Rep gene.
KW slime mould; replication; Rep gene; ss.
OS Dictyostelium discoideum.
FH Key Location/Qualifiers
FT cds 2378..5041
FT /*tag= a
FT /product= involved in extrachromosomal replication
PN WO9106644-A.
PD 16-MAY-1991.
PF 02-NOV-1990; AU0530.
PR 02-NOV-1989; AU-007187.
PA (UYMA-) MACQUARIE UNIV.
PI Slade MB, Chang ACM, Williams KL;
DR WPI; 91-164194/22.
DR P-PSDB; R11988.
PT Polypeptide facilitating extra-chromosomal replication - of
PT recombinant plasmid in Dictyostelium species
PS Claim 15; Fig 1; 90pp; English.
CC The sequence of Ddp2 has been found to contain the putative open
CC reading frame indicated in the Features Table. The possible ORF is
CC flanked by regions with similarity to promoter and poly adenylation
CC signals of known Dictyostelium genes. The RNA and polypeptide
CC product of the Rep gene have not, however, been detected. It is
CC believed that the product is produced in low amounts to positively
CC regulate initiation of plasmid replication. The polypeptide may also
CC contain regions that act as negative regulators of plasmid copy
CC number. See also Q11711 and Q11712.
SQ Sequence 5852 BP; 2298 A; 651 C; 708 G; 2195 T;

Query Match 2.8%; Score 84.4; DB 1; Length 5852;
Best Local Similarity 46.2%; Pred. No. 0.0011;
Matches 317; Conservative 0; Mismatches 366; Indels 3; Gaps 1;

Qy 1997 ttctctttaatcctttcttttacttcatttataacgaattctatggataatgttccct 2056
Db 1635 TTTTTTTTGTGATGACACTTTTTTTTTTGTGATGACACTTTTTTTTAAAAA 1694
Qy 2057 acaaacatgtcattacaagttaattataaattccattcttctattttactaagatatt 2116
Db 1695 AAAAAAATGTTAAATACTATTGTGATGACATTCTTTTCTAGTTTTTTTTAGATAGA 1754
Qy 2117 agtaacttcaaacgtctgatttttactaatttattttataaaattgttagaagtatt 2176
Db 1755 TATAAAATTAATTCGCTATCGATATATCTTAATTATTAAAGATTGAATAATATTITAA 1814
Qy 2177 ttttcaataatttaacacaatatttaattattattattattttctcaattttat 2236
Db 1815 TTTTAAATAATCTACTTTTTTTTTTTTTCTTTTTTTTAAATTTAAATTTTTT 1874
Qy 2237 taacacaaacataaatttttgacaaattaaaataatgaattatttctcaatttttcg 2296
Db 1875 TTTTTATTAGATCTCATATTAATAATCAATTTAAATTAAGATTATTTTAAATATG 1934
Qy 2297 tgcaactattacaaaaatccttcatagtctcaatcttaattgatgcagagtgataata 2356

Db 1935 CAAAACTATAAAAACTAATGTAGTTTAACCACTTTTTTCTATTCTTTTTTTTTT 1994
Qy 2357 atcttaattgatgcagaggttaaatggccgggttgagctggacttaagcatgatat 2416
Db 1995 TTTTTTTTTTACTTTGAAAAA 2054
Qy 2417 tgacgtactttatattttccaaattcaaccagctcgaaatagagtcataaatttgt 2476
Db 2055 ATATTAATTACTTTGGTTT---TTTTGATTTTTTTTAAATAATTTAAATTTTATTC 2111
Qy 2477 ccaatttaattcaagccatttttaagttcgctcattatttttaattaaaaaattta 2536
Db 2112 TCTATCTAATATACCTTATTATAAATATGGAATAATATCAATATTATTCAGTTT 2171
Qy 2537 tatcattttattttaatttaatttttatattttttattttattgaaaattttat 2596
Db 2172 TGGCATGACAATTTAATTATATTATTTTGTAGTTTTTTTTTTTTTTTTTTT 2231
Qy 2597 atagtcattcttaacattatgttaattttattattagtagtattatatatttagtat 2656
Db 2232 AAATTTCTTTTTTTTTTTTTTTTAAATTTTAAATTTTATTTTCCACACTTCA 2291
Qy 2657 aggtttattttgttaataaacttaa 2682
Db 2292 TTTTATTATTATTATTATTGTAA 2317

RESULT 15
Q28302

ID Q28302 standard; DNA; 1511 BP.
AC Q28302;
DT 12-FEB-1993 (first entry)
DE AmEPV tk DNA.
KW Entomopoxvirus; thymidine kinase; non-essential; regulatory sequences;
KW vector; ss.
OS Amsacta moorei.
FH Key Location/Qualifiers
FT cds 852..1511
FT /*tag= a
FT /label= ORF_Q3
FT cds complement (234..782)
FT /*tag= b
FT /label= ORF_Q2
FT CDS complement (17..218)
FT /*tag= c
FT /label= ORF_Q1
FT promoter 750..890
FT /*tag= d
PN WO9214818-A.
PD 03-SEP-1992.
PF 12-FEB-1992; U00855.
PR 19-FEB-1991; US-657584.
PR 30-JAN-1992; US-827685.
PA (UYFL) UNIV FLORIDA.
PI Gruidl ME, Hall RL, Moyer RW;
DR WPI; 92-316172/38.
DR P-PSDB; R29653-55.
PT New viral vectors and chimeric vaccines - comprise entomopoxvirus
PT expression system contg. spheroidin or thymidine kinase sequences
PS Disclosure; Fig 3; 110pp; English.
CC The sequence given is derived from the Entomopoxvirus, Amsacta moorei
CC (AmEPV) and contains the thymidine kinase (tk) DNA sequence. The open
CC reading frames indicated in the features table encode the tk protein
CC itself and also other structural or regulatory genes associated with
CC tk. The tk gene maps near the left end of the physical map of the
CC AmEPV genome. This gene is not highly related to any other
CC vertebrate poxvirus tk gene. Thymidine kinase is a non-essential
CC protein which makes it's gene desirable as a site for the insertion of
CC exogenous DNA.
SQ Sequence 1511 BP; 640 A; 128 C; 98 G; 645 T;

Query Match 2.8%; Score 84.2; DB 1; Length 1511;

Qy	1979	caattttcgagtagtaagtttcotctttaacottctcttttaactcattttataacgaatt	2038
Db	5586	CAATTCCTTT	5527
Qy	2039	ctatggataaatgttccctacaaacatgtcattacaagtgttaattataaaattccattctt	2098
Db	5526	TTT	5467
Qy	2099	ctatttttactaagatattagtaacttcaaacgtctgatttttactaattttattatttata	2158
Db	5466	TTT	5407
Qy	2159	aattgttagaatgattatttttcaataatttaacaacaatattttaattattattattatta	2218
Db	5406	TTT	5347
Qy	2219	ttattttctcaattttttataaacaataaataatttttgacaattaaaataaatgaat	2278
Db	5346	TTT	5287
Qy	2279	taattttctcaatttttcgtgcaactattacaaaatccttcacatgctcctaattcttaatt	2338
Db	5286	TTT	5227
Qy	2339	gatgcagagggtgataataatcttaatttgatgcagaggtaataatgggcgggtttgagc	2398
Db	5226	TTT	5167
Qy	2399	tggacttaagcatgatattgacgtactttatatttttccaaattcaaccagctcgaat	2458
Db	5166	TTT	5107
Qy	2459	atgagctcaaaattttgtccaatttaaccaagcccattttaagttcgccatattatttt	2518
Db	5106	TTT	5047
Qy	2519	tttaattttaaaaatttatatcatcttttttttaattattttatattttttttttta	2578
Db	5046	TTT	4987
Qy	2579	tttattgaaaaatttttatatagtcacattatgtttaatttttatattagagtagt	2638

[illegible]

RESULT 12
X33184/c

ID X33184 standard; DNA; 7996 BP.
AC X33184;
DT 25-JUN-1999 (first entry)
DE Base sequence of the plasmid pRxBcl 2-1-hCD 25.
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9913073-A2.
PD 18-MAR-1999.
PF 07-SEP-1998; J04010.
PR 08-SEP-1997; JP-259235.
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI Hamada H;
DR WPI; 99-243728/20.
PT New apoptosis-resistant virus-sensitive cell
PS Example 3; Page 46-49; 51pp; English.
CC The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the
CC plasmid pRxBcl 2-1-hCD 25, which contains the human Bcl-2 gene, and
CC is used in an example from the present invention.
SO Sequence 7996 BP: 2463 A: 2015 C: 1829 G: 1689 T:

Query Match 3.0%; Score 90.8; DB 1; Length 7996;
Best Local Similarity 45.4%; Pred. No. 0.00017;
Matches 326; Conservative 0; Mismatches 392; Indels 0; Gaps 0;

Qy 1979 caatttttcgagtagtaagtctcctttaacotcttctttttacttcattttataacgaatt 2038
||||| | | | |||| | ||| ||||| | ||||| |
Db 5785 CAATTCCTT 5726

Qy 2039 ctatggataatgttccctacaacatgtcattacaagtgttaattataaaattccattctt 2098
||| | | || | | | | | | | |||| | ||| | | |||
Db 5725 TT 5666

Qy 2099 ctattttactaagaatatttagtaacctcaaacgtctgatttttactaattattattttata 2158
||| || | | || | | | | | | |||| | ||| | |||
Db 5665 TTTTTTTTAGAATGATTATTTTCAATAATTAACAACAAATATTTAATATTATTATTATA 5606

Qy 2159 aattgttagaatgattatttttcaataatttaacaacaatattttaatatattattattata 2218
||| || | || |||| | |||| | |||| | ||||
Db 5605 TTTTGTAGGAATGATTATTTTCAATAATTAACAACAATATTTAATATTATTATTATA 5546

Qy 783 atggaagggaaattgagagtaagttcatgtttattatatacataatgaagttgatgttt 842
Db 300 ATATCAAAATAAACACTTTTATAATAATACGAAAAATATATTCTTATTTTATGTTT 359
Qy 843 tcttcttttataatattttatatacaaaatatttaataaaat----aattaaggattgaat 898
Db 360 TCAAAATTTAGTAGACTTATAATATTATTATGGATAACATTAACAAATAAAATATAT 419
Qy 899 gaaaataataatgaagtcgttttactaatagtcattttgtcgcacatcactta 958
Db 420 GAGTATAATATGTAATTTATTTTATTTTACAGTTTATATGTTTATGAACATATAT 479
Qy 959 aataatagataaataattgtgtacattagatcaaaacaacatagattttgtcccat 1018
Db 480 GTGATAAATAAATTGATTAATTATTATTATATATTAATCTTGTATTTATTAATAA 539
Qy 1019 tctattgtttaaagctggttcattataaaataaggtacattgtacatgccatgt 1078
Db 540 GGTATATTATATATATATATATTTTATTTTATTTTGAATAAAATATTA--AAT 597
Qy 1079 aactatctgttattctatcaatcagcgaatttttaacagtagaagtgatgaatttt 1138
Db 598 AAAAATTTTGTGTTTGGGTAATCATAAAGTCTCAACGTTCAATTTATCTCATTA 657
Qy 1139 taaatagaaaggtcaaaattgttattgatctaacacgtgaggttaattactatttt 1198
Db 658 AAAATAGAAATG--AAATATAATATTACGACAGTACATATATATATATATATATTA 715
Qy 1199 cctaagaagaataagtaaaatataattgaattttaatacaaaaactttcatgatacttt 1258
Db 716 AAAAAAATAAATAAACACAT-ATATATATATATATATATATATATATATATATATAT 774
Qy 1259 atcatattttactataatttaattgtgagagtaacaaarttaaaacacatagaa 1315
Db 775 GTTTTAAAGTATGGATAAATCAAAAAGTCCATAGAGAGAAGAAATTAATAGGATAAAA 831

RESULT 6

T72882

ID T72882 standard; cDNA; 19124 BP.

AC T72882;

DT 12-SEP-1997 (first entry)

DE Plasmodium var-7 gene.

KW DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy;

KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;

KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;

KW Plasmodium; ss.

OS Plasmodium vivax.

OS Plasmodium falciparum.

FH Key Location/Qualifiers

FT exon 7317..15139

FT /*tag= a

FT /number= 1

FT intron 15140..16205

FT /*tag= b

FT /number= 1

FT exon 16206..17552

FT /*tag= c

FT /number= 2

FT /note= "no stop codon given"

PN WO9640766-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; U09508.

PR 07-JUN-1995; US-487826.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

PI Welles TE;

DR WPI; 97-052231/05.

DR P-PSDB; W22475.

PT New malaria vaccines - contains cysteine-rich DBL family protein

PT binding domains homologous domains of the Duffy and sialic acid

PT binding proteins

PS Claim 4; Page 56-61; 96pp; English.

CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the invention.
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).

SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;

Query Match 3.2%; Score 96.2; DB 1; Length 19124;
Best Local Similarity 45.1%; Pred. No. 3.2e-05;
Matches 591; Conservative 0; Mismatches 709; Indels 10; Gaps 6;

Qy 138 aatataataacatcgtagaataaattttattcaaaattgaagtc---taaccatct 194
Db 5418 AATACGTACATGTTATATAGAAATAAAGAATTTAATATTAAGGATAAATAAATAT 5477
Qy 195 ttaattttgtatgtatatttaaatgaagataaatacatattcttgacatgtatttt 254
Db 5478 TTAATAATTATATTTTATGTCAATTTATGTATATATATATATTAACATGATAGT 5537
Qy 255 catcttaattgttggcttggtagatgtattgtatgtacgatgtcttttaaacac 314
Db 5538 TTTTGAATAATTTAAATATCATATAATAATAAATAGTTAAATAATAGTATTT 5597
Qy 315 atatcacattttgagttgtatgatgataagtcacataancgaaataggtgtatctt 374
Db 5598 CATACAAATACTAACTTATAAGTATATCATATAAATATATATATATATATTATGTG 5657
Qy 375 cacttttgaactttgataagtcacaaactttaacaagtttg---attgtgtacata 431
Db 5658 TTTTGATTGGGTGATATAAGGCTAAGTATATATGGGTGTTTCATTATATATTATA 5717
Qy 432 tatatatatttcaattttataataaaattgttttaataatttacagttatattat 491
Db 5718 TGTGAATAGATACATATAAGTTAAT-ATATTATTGTGTATATGTCGTGTGAAGATAG 5776
Qy 492 tttttatctctaatttttattgtcgcaaatttttagttgatatttaacataaaaaa 551
Db 5777 ATATGATTACAGTTAAGGGTTATAGTTTTTTTTTTTTTTTGTACATATATATAAA 5836
Qy 552 attgtacacatttacaagccatatacaataattataaatattcattaaaaatata 611
Db 5837 AATAGATAACTAACAATATGCATATTACAAGAATAAATTTGTATATAAAT-ATATATA 5895
Qy 612 ttttaatataggatataaataaactattttagattttctacttttaagataacatagg 671
Db 5896 TATATATATAAAGACATTAAACATATACTAATAAGGTAATAGTTTATTATATCATCC 5955
Qy 672 ttaattgtataaataaagggttagttttgttaagtagatgatattgtcgtaaacat 731
Db 5956 TTTTATTATATAAATTTTTTTTCTTCTGTCGTTCTTTTGTATTATATAATAT 6015
Qy 732 aatcactaacctttttattaaactttgtgtttgaagttccaaaaaagaatggaagg 791
Db 6016 AACAAATATAAACAATATCAGTATTTGGAAATATAAATAAATTTATTCTCATATATGCA 6075
Qy 792 aaatttgagagtaagttcatgtttattatatacataatgaagttgatgttttctttt 851

Db 181 AGTCTTAACCATCTTTAATATTGTAGATGAATTAATGAAGATAAATACATATTCT 240
Qy 241 tggacatgtatcttcatcttaagtgttggtggtgataggtgattgatgcgat 300
Db 241 TGGACATGTATTTTCATCTTAATGTTTGTGGCTTGGTGATAGGTGATGATGACGAT 300
Qy 301 gtcttttaaatcacatatcacattttgagtttgatgatgaagtcacataancgaaa 360
Db 301 GTCTTTAAATCACATATCACATTTTGAGTTTGATGATGATAGTCGACATAANCGAAA 360
Qy 361 tatggtgtgatcttcacttttgaaacttgataagtcacaaactttaacaaagttgatt 420
Db 361 TATGGTGTGATCTTCATCTTTGAACTTTGATAGTCACCAAACTTTAACAAGTTGATT 420
Qy 421 gtgtacatatatatatatcttcaaattttataataaaaattgtgttaataattac 480
Db 421 GTGTACATATATATATATATCTTCAAAATTTATAATAAAAAATGTGTTAAATATTAC 480
Qy 481 agttatattatcttct 540
Db 481 AGTTATATTATTTTATCTCTAATTTATTTGTCGCAAAATTTTAGTTGATATTTTA 540
Qy 541 acataaaaaaattgtacacatttacaagccatatacaataattatataaattcat 600
Db 541 ACATAAAAAAATTGTACACATTTACAAGCCATATACAAATTAATATATAAATTCAT 600
Qy 601 taaaaatatatttaaatataggatataaataactattttagaattattctacttaa 660
Db 601 TAAAAATATATTAAATATAGGATATAAATAAATAAATTTAGAATTATCTACTTTAA 660
Qy 661 gataacataggttaaatgtataaataaaggtagtttttgaagatgagtatatat 720
Db 661 GATAACATAGGTAAATGTATAAATAAAGGTAGTTTATGTAAGATGAGTATATAT 720
Qy 721 gtctgaacataatcactaacctttttataactctctgttttgaagttccaaaaaga 780
Db 721 GTCGTAACATAATCACTAACCTTTTATTAACTCTTGGTTTGAAGTCCAAAAAGA 780
Qy 781 aaatggaagggaatttgagagtaagttcatgtttatattatacaataatgaagttgat 840
Db 781 AAATGGAAGGGAATTGAGAGTAAGTTTATATTATATACATAATGAAGTTGATGT 840
Qy 841 tttctcttttaattttttatacaaaaattttaaaaaaataaaggattgaatga 900
Db 841 TTTCTCTTTTAAATTTTATATACAAAATTTTAAATAAAATAATTAGGATTGAATGA 900
Qy 901 aaaaataatgaagtcggttttactaatagtcataattgttcgtcatctacttaa 960
Db 901 AAAATATAATGAAGTCGTTTACTAATAGTCATATTGCAATTTTGTGCATCTACTTAA 960
Qy 961 taatagataaaatattgtggtacattagatcaaagaacaaactagattttgtccattc 1020
Db 961 TAATAGATAAATAATTGTGGTACATTAGATCAAGAAACAACTAGATTTTGTCCATTTC 1020
Qy 1021 tattgttaaaagctggtccgtttacataaaataaggtacatgttacatgccagtaaa 1080
Db 1021 TATTGTAAAGCTGGTCCGTTTACATTAATAAGGTACATGTTACATGCCAGTATAA 1080
Qy 1081 ctatctggttattctcaatcacgctaatttttaacagtagaagtgaaattttta 1140
Db 1081 CTATCTGGTTATTCTCAATCACGCTAATTTTAACTAGAAATGAATGTAATTTTA 1140
Qy 1141 aatagaagggtcaaatgttattgtatcaacagtagggttaattactattttcc 1200
Db 1141 AATAGAAGGGTCAAAATGTTATTGTATCAACAGTAGGATTAATTACTTATTTTCC 1200
Qy 1201 taaagaataagtaaaatataatttgaatcttaatacaaaaactttcatgatactttat 1260
Db 1201 TAAAGAATAAGTAAATATAATTGAATCTTAATACAAAACCTTCATGATACTTTAT 1260
Qy 1261 catattttactataatttaattgtgagagtaacaaarttaaaaaacatagaacacc 1320

Db 1261 CATATTTTACTTATAATTTAATATTGTGAGAGTAACAAARTTAAAAACATAGAACACC 1320
Qy 1321 aaaagttagttatggtgtgactcatatacacagttaaaattgaataaattttttcttc 1380
Db 1321 AAAAGTTAGTTATGGTGTGACTCATATACACAGTTAAAATTTGAATAAATTTTCTTTC 1380
Qy 1381 gtcattaattccatcatgggttttttttttctagtttaagccataattatcaaaataatc 1440
Db 1381 GTCATTAATCCATCATGGGTTTTTTTTTCTAGTTAAGCCATAATTATCAAAATAATC 1440
Qy 1441 atcattaatcctatcaataccccgcctgctccctccctcaactactaaacccaactaa 1500
Db 1441 ATCATTAACTCATCAATACCCGCCCTGCTCCCTCCCTCAACTTAAACCAACTAA 1500
Qy 1501 caccagcagcaaacgcactttaatagccactattttctagccatgtcttgcacttaa 1560
Db 1501 CACCAGCAGCAAAACGCACTTTAATAGCCACTATTTCAGCCATGCTTGCACCTAAA 1560
Qy 1561 gaaaagtaagctaacctgcaatcattccatcgaggcctcaacagataaagttggtg 1620
Db 1561 GAAAAGTAAGCTAACCTGCAATCATTCCATATCGAGGCTCAACAGATAAAGTTGGTGG 1620
Qy 1621 atgggtttgcaccaagttgttaaaacccgcctcaacttccctttttctctccctcc 1680
Db 1621 ATGGGTTGCAACAGTTGTAAAACCCGCCCTCAACTTCCCTTTTCTTTTCATCTCC 1680
Qy 1681 ccactccacccctccaattttctctatgttctattataaagttctttataatcacag 1740
Db 1681 CCACTCCACCCCTCCAATTTCTTCATATGGTTCTATTATAAGTTCTTTATAATCACAG 1740
Qy 1741 aatcaagataagctcagcaaaaaaaccaatggctcgcagcaagatctgactagt 1800
Db 1741 AATCAAGATAAGCTCAGCAAAACAAAACCATGGCTCGAGCAAGATCTGACTAGT 1800
Qy 1801 cagagctctgaattattggtacattattacagtcacaaacagttacaaaagctgttgcag 1860
Db 1801 CAGAGCTCTGAATATTGGATCATTATTACAGTCAAAACAGTTAACAAGCTGTTGCAG 1860
Qy 1861 ataaacactgaatctgctatagttgtttttgttttacatagttccagtgaaactatg 1920
Db 1861 ATAAACACTGAATCTGCTATAGTTGTTTTGTGTTTACATATGTTCCAGTGAACATG 1920
Qy 1921 aagcatctctaagaaaacccaactatcatatcaacccatcgatcaatgaatcgatttca 1980
Db 1921 AAGCATCTCTAAGAAAACCAACTATCATATCAACCCATCGATCAATGAATCGATTTC 1980
Qy 1981 attttgcagataagttctcttttaactcttttttacttattttataacgaattct 2040
Db 1981 ATTTTGCAGTATAAGTTCTTTTAACTCTTTTACTTCTTTTATAACGAATTCT 2040
Qy 2041 atggataatgttccctacaaacatgctattacaatgtttaattataaattccattctct 2100
Db 2041 ATGGATAATGTTCCCTACAAACATGCTATTACAATGTTAATTATAAATTCATCTCT 2100
Qy 2101 attttactaagatattagtaactcaaacctgctgatttttactaattattattataaa 2160
Db 2101 ATTTTACTAAGATATTAGTAACCTCAAACTGCTGATTTTACTAATTATTATTATAAA 2160
Qy 2161 ttgttagaattgattttttcaataatttaacacaatttttaattattattattatt 2220
Db 2161 TTGTTAGAATGATTATTTTCAATAATTAAACAATAATTATAATTATTATTATTATT 2220
Qy 2221 attttcattttttataaacaacaaataaattttgacaaattaaaaataaataa 2280
Db 2221 ATTTCTCAATTTTATAACAAAACATAAATTTTGACAAATTAATAAATGAATTA 2280
Qy 2281 attttcattttttctgcaactattacaaatccttcatagtcctaatcttaatttga 2340
Db 2281 ATTTCTCAATTTTCTGCAACTATTACAAAATCCTTCATAGTCTTAATCTTAATTGA 2340
Qy 2341 tgcagaggtgataaataatcttaattgtgagaggttaaatggcggtttgagctg 2400
Db 2341 TGCAGAGGTGATAAATCTTAATTGTGAGAGGTAAATGGCGGGTTTGAGCTG 2400

Db 139996 TTCATCTATATATAATCATATTAATTATATTCATTCTATATATATATCTAATCA 140055

Qy 2859 attgagcttaattaatt 2877

Db 140056 ATTAATAACATATAATAAT 140074

RESULT 15

AC005504
LOCUS AC005504 104992 bp DNA HTG 01-APR-1999
DEFINITION Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
ACCESSION AC005504
VERSION AC005504.3 GI:4558584
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 104992)
AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 104992)
AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT On Apr 2, 1999 this sequence version replaced gi:4337172.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 58642: contig of 58642 bp in length
* 58643 58842: gap of unknown length
* 58843 91011: contig of 32169 bp in length
* 91012 91211: gap of unknown length
* 91212 104992: contig of 13781 bp in length.
FEATURES Location/Qualifiers
source 1. 104992
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
BASE COUNT 44286 a 9326 c 9564 g 41411 t 405 others
ORIGIN

Query Match 5.1%; Score 156.6; DB 41; Length 104992;
Best Local Similarity 44.6%; Pred. No. 6.3e-09;
Matches 1232; Conservative 0; Mismatches 1491; Indels 42; Gaps 14;

Qy 22 atagtaaacctaaccaatttttaataaagctgactcctagtacaagagctttttatt 81

Db 72473 ATAAATACATATTAATATATATATATTTATTAATTTATAAAATAAATAATATA 72532

Qy 82 cattctctattttgtcttctcttaggcttggaatcgagaattttctgtgtacaata 141

Db 72533 TATATATAATTATATAATATATCAAAATTAATCATTATAAAATTTATTTAAATATATTA 72592

Qy 142 taataaatatcatgtagaataaattttattcaattgaagctcttaaccatctttaatat 201

Db 72593 AAATTAA-ATATATATATTAATAAATAAATTAAGTTAATTTATTAATAAATAAATAA 72651

Qy 202 ttgtagatgtaatttaataagataaatacatattcttgacatgtattttcatotta 261

Db 72652 TAATAAATTTAAATTAATATAATTAATTCATAATACACATTAATTAATAAATATGA 72711

Qy 262 atgtttgtggttggtgataggtgtattgtatcagatgtcttttaaacacatatcac 321

Db 72712 ATATTATATAATAATAATAAGAAAAATATTAAACAAATTTAAATATTAAATAATAA 72771

Qy 322 attttgagttgtatgatgaagtcagacataancgaatatgtgtgatcttcactttt 381

Db 72772 AATATTATAATTTATAATAATAAATAATTAATAATAAATTAATTAATAATATAATAA 72831

Qy 382 gaactttgataagtcaccaaactttaacaaagtttgattgtgtacatatatatatc 441

Db 72832 TTAATATAAATTAATTTAAATATAAATTAATAAATAAATAATACTAATATTAAATAA 72891

Qy 442 ttcaaatttttataaaaaattgtgtttaaataattacagttatatttttttatct 501

Db 72892 ATAAATAATAATAATAAATTTAAATTAATAAATAAATAATAAATAATAATAA 72947

Qy 502 ctaattttttgtgcgaatatttttagttgatattttaacataaaaaaattgtacaca 561

Db 72948 ATTAATTAATAATAATAATAAATAAATAAATAATAATAATAATAATAATAATA 73007

Qy 562 ttacaagcccatatacaataattataataaattacataaaaaatatatttaataata 621

Db 73008 ACAATATTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73067

Qy 622 ggatataaataaactatttttagaattattctactttaagataacataggttaagtgt 681

Db 73068 AGAATTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73117

Qy 682 aattaataaggttagttttttgtaagatgagtatatgtcgtaaacataatcactaac 741

Db 73118 AATAATAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73173

Qy 742 catttttttaacttctgtgtttggaagttccaaaagaaaattggaaggaattgaga 801

Db 73174 ATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73233

Qy 802 gtaagttcatgtttatattatacaataaagttgatgttttcttcttttaataatttt 861

Db 73234 ATATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73292

Qy 862 atacaaaattatttaataaataaataaagattgaatgaaaataatgaaagtcgttt 921

Db 73293 ATATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73352

Qy 922 tactaatagtcattgtcattttgtcgcatctacttaataatagataaattattgtgg 981

Db 73353 AATTAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73412

Qy 982 tacattagatcaaagaacaaactagattttgtccattctattgttaaaagctggtccgt 1041

Db 73413 TTAATTAATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73472

Qy 1042 ttacat-taaaaaagggtacatgttacatgcacgtataactatctgtttattctcaa 1100

Db 73473 TAATATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73532

Qy 1101 tcacgttaatttttaacagtagaataatgaatgtaatttttaataagaaaggtcaattgt 1160

Db 73533 ACAATAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73592

Qy 1161 tattgtatctaacagtaggattaatttactatttttctaaagaaataagtaaaatat 1220

Db 73593 TAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73649

Qy 1221 aattggaatcttaatacaaaaaactttcatgacttttatcatattttacttataatta 1280

Db 73650 AATATACATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73709

Qy 1281 atattgtgagagtaacaaarttaaaaaacatagaaacacaaaggttagttgtgtga 1340

Db 73710 AATTTAAGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 73769

Qy 1341 ctcatatcacagttaaaaattgtaataattttttcttcgtcattaattccatcatggg 1400

* 10240 10319: gap of unknown length
* 10320 10964: contig of 645 bp in length
* 10965 11044: gap of unknown length
* 11045 11648: contig of 604 bp in length
* 11649 11728: gap of unknown length
* 11729 12696: contig of 968 bp in length
* 12697 12776: gap of unknown length
* 12777 13976: contig of 1200 bp in length
* 13977 14056: gap of unknown length
* 14057 15045: contig of 989 bp in length
* 15046 15125: gap of unknown length
* 15126 15969: contig of 844 bp in length
* 15970 16049: gap of unknown length
* 16050 16859: contig of 810 bp in length
* 16860 16939: gap of unknown length
* 16940 17662: contig of 723 bp in length
* 17663 17742: gap of unknown length
* 17743 18767: contig of 1025 bp in length
* 18768 18847: gap of unknown length
* 18848 19809: contig of 962 bp in length
* 19810 19889: gap of unknown length
* 19890 21046: contig of 1157 bp in length
* 21047 21126: gap of unknown length
* 21127 21826: contig of 700 bp in length
* 21827 21906: gap of unknown length
* 21907 23735: contig of 1829 bp in length
* 23736 23815: gap of unknown length
* 23816 25556: contig of 1741 bp in length
* 25557 25636: gap of unknown length
* 25637 26792: contig of 1156 bp in length
* 26793 26872: gap of unknown length
* 26873 28359: contig of 1487 bp in length
* 28360 28439: gap of unknown length
* 28440 29898: contig of 1459 bp in length
* 29899 29978: gap of unknown length
* 29979 31836: contig of 1858 bp in length
* 31837 31916: gap of unknown length
* 31917 33347: contig of 1431 bp in length
* 33348 33427: gap of unknown length
* 33428 34568: contig of 1141 bp in length
* 34569 34648: gap of unknown length
* 34649 35754: contig of 1106 bp in length
* 35755 35834: gap of unknown length
* 35835 37815: contig of 1981 bp in length
* 37816 37895: gap of unknown length
* 37896 39641: contig of 1746 bp in length
* 39642 39721: gap of unknown length
* 39722 41135: contig of 1414 bp in length
* 41136 41215: gap of unknown length
* 41216 42477: contig of 1262 bp in length
* 42478 42557: gap of unknown length
* 42558 44229: contig of 1672 bp in length
* 44230 44309: gap of unknown length
* 44310 45922: contig of 1613 bp in length
* 45923 46002: gap of unknown length
* 46003 47999: contig of 1997 bp in length
* 48000 48079: gap of unknown length
* 48080 49982: contig of 1903 bp in length
* 49983 50062: gap of unknown length
* 50063 51360: contig of 1298 bp in length
* 51361 51440: gap of unknown length
* 51441 53101: contig of 1661 bp in length
* 53102 53181: gap of unknown length
* 53182 54926: contig of 1745 bp in length
* 54927 55006: gap of unknown length
* 55007 56937: contig of 1931 bp in length
* 56938 57017: gap of unknown length
* 57018 57606: contig of 589 bp in length
* 57607 57686: gap of unknown length
* 57687 58632: contig of 946 bp in length
* 58633 58712: gap of unknown length
* 58713 60613: contig of 1901 bp in length
* 60614 60693: gap of unknown length

* 60694 62727: contig of 2034 bp in length
* 62728 62807: gap of unknown length
* 62808 65311: contig of 2504 bp in length
* 65312 65391: gap of unknown length
* 65392 66685: contig of 1294 bp in length
* 66686 66765: gap of unknown length
* 66766 68830: contig of 2065 bp in length
* 68831 68910: gap of unknown length
* 68911 71103: contig of 2193 bp in length
* 71104 71183: gap of unknown length
* 71184 72193: contig of 1010 bp in length
* 72194 72273: gap of unknown length
* 72274 74138: contig of 1865 bp in length
* 74139 74218: gap of unknown length
* 74219 76236: contig of 2018 bp in length
* 76237 76316: gap of unknown length
* 76317 77913: contig of 1597 bp in length
* 77914 77993: gap of unknown length
* 77994 80808: contig of 2815 bp in length
* 80809 80888: gap of unknown length
* 80889 82776: contig of 1888 bp in length
* 82777 82856: gap of unknown length
* 82857 85682: contig of 2826 bp in length
* 85683 85762: gap of unknown length
* 85763 89308: contig of 3546 bp in length
* 89309 89388: gap of unknown length
* 89389 93999: contig of 4611 bp in length
* 94000 94079: gap of unknown length
* 94080 109466: contig of 15387 bp in length
* 109467 109546: gap of unknown length
* 109547 110183: contig of 637 bp in length
* 110184 110263: gap of unknown length
* 110264 110842: contig of 579 bp in length
* 110843 110922: gap of unknown length
* 110923 111636: contig of 714 bp in length
* 111637 111716: gap of unknown length
* 111717 112030: contig of 314 bp in length
* 112031 112110: gap of unknown length
* 112111 112674: contig of 564 bp in length
* 112675 112754: gap of unknown length
* 112755 113301: contig of 547 bp in length
* 113302 113381: gap of unknown length
* 113382 113979: contig of 598 bp in length
* 113980 114059: gap of unknown length
* 114060 114698: contig of 639 bp in length
* 114699 114778: gap of unknown length
* 114779 115187: contig of 409 bp in length
* 115188 115267: gap of unknown length
* 115268 115938: contig of 671 bp in length
* 115939 116018: gap of unknown length
* 116019 116539: contig of 521 bp in length
* 116540 116619: gap of unknown length

Query Match 5.3%; Score 162.6; DB 55; Length 161891;
Best Local Similarity 41.4%; Pred. No. 1.3e-09;
Matches 1027; Conservative 1; Mismatches 1425; Indels 26; Gaps 8;

Qy 411 aagtttgattgtgtacatatatatatatcttcaaaattttataaaaaattgtgttta 470
||| ||| || |||| ||| |||| |||| ||||
Db 137610 AAGGATGTTATATTAAATATACCGACATTAATAATAATACCAATATTTTATTGTTAT 137669

Qy 471 aataatttcagttatattttttttatctctatattttttgtcgccaaatttttagt 530
| | | | | | | | | | | | | | | | | | | |
Db 137670 ATAATATAAATCAAAATATTTTAAACACATATATAAATATAAATTTTATATTTAATCAT 137729

Qy 531 tgatatttttaacataaaaaaaattgtcacacattaccagcccatatacaataattatat 590
| | | | | | | | | | | | | | | | | | | |
Db 137730 ATAAATAAAATATTTTAACTTAATATATTTTATATAAATTTTAAATTTTAAATAT 137789

Qy 591 aaatattcattaaaaaatattttaaatat---aggatataaaataactattttagaat 647
|| |||| ||| || |||| |||| |||| |||| ||||
Db 137790 AATTATTTTATTTTATTTATATTACATATTAATTTATATAAATTTTAAATTTTAAATATA 137849

Qy 127 tcttggtgtaacaataataataacatcgtagaaataaattttattcaaatgaagtctt 186
| | | | | | | | | | | | | | | | | |
Db 98689 TATTTTCGTAATAAATAATAATAAAATGATATTAAATATTATATTAACTACAATAT 98748

Qy 187 aaccatctttaatatattgtagatgtaatttaaagtgaagaataaacatattcctggaca 246
| | | | | | | | | | | | | | | | | |
Db 98749 TAATATTTTTATTATTATAAAATATTATTAAAAATATATAAATATTAAATTTATGGAAT 98808

Qy 247 tgtattttcatcttaagtgttggcttgggtgatagggtattgtatgcagtgtcttt 306
| | | | | | | | | | | | | | | | | |
Db 98809 TTTTAAATTAATTTAAATTAATGTCTTATATTAAATTATATATTTA-ATAAAATATATTT 98867

Qy 307 taaatcacatatcacattttgagtttgtatgatgataagtcgcacataancgaaatattggt 366
| | | | | | | | | | | | | | | | | |
Db 98868 TAAAATAATACACACAAAATGATTCTTAATTAATATAAAATATTTATTTTATTATAAT 98927

Qy 367 gtgatottcacttttgaaccttgataagtcaccaaacctttaacaaaagtttgattgtgtac 426
| | | | | | | | | | | | | | | | | |
Db 98928 GAAATAAATAATTTAATTTAAATAAAATAAAATAAAATAATAATATTATTATTATAT 98987

Qy 427 atatatatatatatcttcaaattttataataaaaaattgtgtttaaataatttcacgittat 486
| | | | | | | | | | | | | | | | | |
Db 98988 A-ATATATTTAATTAATTAATTAATTTATTTAATTAATTTAAATAAAATAAAATAAT 99046

Qy 487 attatttttttatctctaatttttattgtcgccaaatttttagttgatatttttaacataa 546
| | | | | | | | | | | | | | | | | |
Db 99047 AAATATTATTATTATATATAATATTTAATTAATTAATTTAATTTAATTTAATTAATTAAT 99106

Qy 547 aaaaaattgtcacacatttacagcccacatatacaaaataattatataaattattcattaaaa 606
| | | | | | | | | | | | | | | | | |
Db 99107 TAAAATAAAATAAAATAAATAATTTATTATTATATAATATATTTAATTAATTAATTAAT 99166

Qy 607 atatattttaaatatagगतataaaataaactattttagaattattctacttaagataac 666
| | | | | | | | | | | | | | | | | |
Db 99167 TATTATTTAATTA-----ATTAAAAATAAATAAATAAATAAATAATATATATATTA 99222

Qy 667 ataggttaaatgtataattaataaggtttagtatttgtaaagtatgatatatgtcgta 726
|| || | ||| | | | | | | | | |
Db 99223 TTAATAAATAAAATATAATTCAATTATATAAATTTTATAAATTTTAAATAAATAATAAAA 99282

Qy 727 aacataatcactaacccctttttataacttcctggtttgaagttccaanaagaaatgg 786
| | | | | | | | | | | | | | | | |
Db 99283 TAATAATAAATAAATAATTTAATAA-----TATAATAACATTATAAATAAATAAAT 99337

Qy 787 aagggaatttgagagtaagttcatgtttatattacataatgaagttgatgtttctt 846
| | | | | | | | | | | | | | | | |
Db 99338 AATTTAATAAATAATATATATAAATTTAATAAATTTAATAACATATATTTTATAATAA 99397

Qy 847 ctttttaattttttatcaaaaattttaataaaaaataaaggattgaatgaaaaata 906
| | | | | | | | | | | | | | | | |
Db 99398 TTTCTTAATTTTTTTTATCATATTATATAATATATATTTTTTTTAAAATAAATAAT 99457

Qy 907 taatgaagtcgttttactaatagtcattgtcatttgtgcgcactctaataatag 966
| | | | | | | | | | | | | | | | |
Db 99458 TAAATAAATAATTTAATAAATAAATAATTTAATAAGTAATAAATAATAACAATAAT 99517

Qy 967 ataataaatttggttacattagatcaagaacaactagattttgtcccattctattgt 1026
| | | | | | | | | | | | | | | | |
Db 99518 ATAAATATATATATATATTTAATAAATAAATAAATAAATAATTTTATATATATATA 99577

Qy 1027 taaaagctggtccgtttaccataaaaagggtacatgtttacatgccaggtataactatc 1086
| | | | | | | | | | | | | | | | |
Db 99578 TATATATTAATAATTAATTTAAATATAAATAAAGAATAATTTTATACTTTTATTA 99637

Qy 1087 ggttatctcatcaatcacgctaatttttaacagtagaagtgaatttttaataaga 1146
| | | | | | | | | | | | | | | | |
Db 99638 ATTAATATATAGTAATAAATAATTTTATGTTATTATATATAATATTTTATTTTAT 99697

Qy 1147 aagggtcaaattgttatttgatctaacacgtagggttaatttactattttcctaaga 1206
| | | | | | | | | | | | | | | | |
Db 99698 TTTTATTATTATAAATAAATAATTTTATAAATAATTTTAAATAAATAAATA 99757

Qy 1207 aataagtaaaatataatttgatcttaatacaaaaactttcatgacctttatcatatt 1266
| | | | | | | | | | | | | | | | |
Db 99758 AACATATAAATAATTAATAAATAATATATATTTTTTTTAAATAAATAAATAATAT 99817

Qy 1267 ttacttataatttaattgtgagagtaacaaartta-----aaaacatagaaacacc 1320
| | | | | | | | | | | | | | | | |
Db 99818 TCACATTTTAATAAATAAATAAACATTATTATAAATACTAATAATAATAAATA 99877

Qy 1321 aaaagttagtttggtgactcatatacacaggtaaaattgaataaattttttcttc 1380
| | | | | | | | | | | | | | | | |
Db 99878 AAAATAATTTAATTGTGAATTAATAAATAAATAAATACTTTAATAATAATATAT 99937

Qy 1381 gtcattaatccatcatgggttttttttctagtttaagccataattatcaaaataatc 1440
| | | | | | | | | | | | | | | | |
Db 99938 ATAATATATATATTATCTTAAATTAATTAATTTTTTAATTTTAAATAATAATA 99997

Qy 1441 atcataatcctatcaataccccgcctgcctccctcctaacttaaacccaactaa 1500
| | | | | | | | | | | | | | | | |
Db 99998 TTTTATTAATTAATTATGTATATTTTATTAATGTTAATTTTATTTATTTTAT 100057

Qy 1501 caccagcaccacaaagcacttttaagccacctatttctagccatgacctgcaactaa 1560
| | | | | | | | | | | | | | | | |
Db 100058 TAATATATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 100117

Qy 1561 gaaaagtaagactaacctgcaatcattccatatcgaggcctcaacagataaagttggtg 1620
| | | | | | | | | | | | | | | | |
Db 100118 ATATTATTAATTTAATTTGTTTATTATTATATATATATTTAATTTAA--TTAATA 100175

Qy 1621 atgggtttgcaccaagttgttaaaaacccggccctcaactccctttttctcatcc 1680
| | | | | | | | | | | | | | | | |
Db 100176 TTTTATTATTATATATATATTATTAATTTATATAAATAAATAATTAATTAATTAAT 100235

Qy 1681 ccactccacacccccaattttcttcacatggttctattataagttctttataatcacag 1740
| | | | | | | | | | | | | | | | |
Db 100236 TAATTTATTATATATTAATTAATTAATTAATTTAATTTAATTTAATTAATTTAAT 100295


```

RESULT 12
AC005504/c
LOCUS AC005504 104992 bp DNA HTG 01-APR-1999
DEFINITION Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
ACCESSION AC005504
VERSION AC005504.3 GI:4558584
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 104992)
AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 104992)
AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Direct Submission

```

Oy 727 aacataatccactaacccatttttattaaacctcttggtttgaagttccaaaaagaaaatgg 786
| | | | | | | | | | | | | | | | | |
Db 74389 TAATAATAAATTAATAATTTAATTAA-----TATAAATAAACATTATAAATTAAGAAAAT 74395

Db 18676 ATATATATATAAATATTTAATATATATTATTATATATCTAATAATTTAAATAAAAAATTTTA 18735

Qy 2577 tattttattgaaaattttttatatagtcatctttaacattatgttaagtgtttatattagagta 2636
| | | | | | | | | | | | | | | | | |

Db 18736 AAAATTTAAAAATGTAGATATAAATTTATAAAAAATTTATATTCTCATATTTATTATTATTATTA 18795

Qy 2637 gtattatataatatttagtataggtttattttgttaataaaccttaaaaaatgggtcttggtg 2596
| | | | | | | | | | | | | | | | | |

Db 18796 ATTTAATTTTATATAAAATAATATAATGATTTAATTAATTA--TTATATATTTTATAAATTTA 18853

Qy 2697 gctagacttggagccttaaatgctcaaacctcaaaccttaattcatattttaaacaggcgttaa 2756
| | | | | | | | | | | | | | | | | |

Db 18854 TATATTATTGAATATTTATATAATATATATATATATAGAAAAATTAATATTTTAAAT 18913

Qy 2757 tattttttattcacactgtttcaaatttttgcgggtgaatatcttcgagctgtagaataa 2816
| | | | | | | | | | | | | | | | | |

Db 18914 AATTTAATATAAATTTTTTAAAAAATTTCTTAAATGTATTATTTTATAAAAAATATTTTAT 18973

Qy 2817 acaccacagggtctaattttagtgcctaatgaaaatgaaatcatattgagcttaattaatat 2876
| | | | | | | | | | | | | | | | | |

Db 18974 ATAATAAAATCATTTTTTTTTAAAAATAAACAAAAAATTTTTTAATAAATAAATTTTATAA 19033

Qy 2877 tccattcttcttttgcgtga 2894
| | | | | | | |

Db 19034 TGAATATATAATTTTATTTA 19051

RESULT 11
AE001398/c
LOCUS AE001398 14867 bp DNA INV 06-NOV-1998
DEFINITION Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.
ACCESSION AE001398 AE001362
VERSION AE001398.1 GI:3845197
KEYWORDS
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium..
REFERENCE 1 (bases 1 to 14867)
AUTHORS Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Perlea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Science 282 (5391), 1126-1132 (1998)
MEDLINE 99021743
REMARK Erratum: [[published erratum appears in Science 1998 Dec 4;282(5395):1827]]
REFERENCE 2 (bases 1 to 14867)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA
FEATURES
source Location/Qualifiers
1..14867
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
gene complement(1570..2424)
/gene="PFB0490c"
CDS complement(1570..2424)
/gene="PFB0490c"
/note="predicted by GlimmerM"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC71887.1"
/db_xref="GI:3845198"
/translation="MKEKNEKIMDYLSCLDDVVDREKSGKGNLLKSSSTKSKSDYKKSSIFSKKRDSHKKGSSFRGRRSGFINRKSFSFKKPYNNRLINKNNYNNYKGRNFHNGRDNFKGRGTSGFSGRVFDNRKGSFKKRFISNRNKS SVKSYRGNGSKNMGRKSFNKAPTSRTVTVKRLNNYKTVSAPVKKFNNLNLSLYRKNRTFALNTRKSPVGTIKSSVPRKRIK

REFERENCE 1 (bases 12511 to 12682)
AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and Wolstenholme,D.R.
TITLE Drosophila mitochondrial DNA: a novel gene order
JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE 83090428

REFERENCE 2 (bases 5269 to 5695)
AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes
JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE 83220794

REFERENCE 3 (bases 404 to 5272)
AUTHORS de Bruijn,M.H.
TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code
JOURNAL Nature 304 (5923), 234-241 (1983)
MEDLINE 83245048

REFERENCE 4 (bases 804 to 1778)
AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.
TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species
JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE 88174373

REFERENCE 5 (bases 5268 to 13619)
AUTHORS Garesse,R.
TITLE Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations
JOURNAL Genetics 118 (4), 649-663 (1988)
MEDLINE 88212147

REFERENCE 6 (bases 441 to 2967)
AUTHORS Satta,Y. and Takahata,N.
TITLE Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE 91088557

REFERENCE 7 (bases 14215 to 14512)
AUTHORS Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.
TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods
JOURNAL Science 258 (5086), 1345-1348 (1992)
MEDLINE 93088057

REFERENCE 8 (bases 14917 to 19517)
AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE 94285822

REFERENCE 9 (bases 1 to 408; 13319 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons
JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE 96423163

REFERENCE 10 (bases 1 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department, Michigan State University, East Lansing, MI 48824-1319, USA

FEATURES
source
1. .19517
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/db_xref="taxon:7227"
/note="derived from new and previously submitted sequences; sequence is a composite containing sequences obtained from different Drosophila melanogaster strains"

tRNA
1. .65
/gene="mt:ND6"
/product="tRNA-Ile"
/db_xref="FlyBase:FBgn0013685"

gene
1. .19517
/gene="mt:ND6"
/note="mitochondrial NADH-ubiquinone oxidoreductase chain 6"
/allele=""
/db_xref="FlyBase:FBgn0013685"

tRNA
complement(97. .165)
/product="tRNA-Gln"

tRNA
171. .239
/gene="mt:ND6"
/product="tRNA-Phe"
/db_xref="FlyBase:FBgn0013685"

CDS
240. .1265
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC47811.1"
/db_xref="GI:1166530"
/translation="MFNNSSKILFITIMIIGTLITVTSNWLGAWMGLEINLLSFIPL
LSDNNNLMSTEASLKYPLTQVLASTVLLFSSILLMLKNNWNEINESFTSMIIMSALL
LKSGAAPFFHFWFNNMSEGLTMMNMLMTWQKIAPMLISYLNKYLILSVILSVII
GAIGGLNQTSRLKLMAPSSINELGWLSSLMISEINLWLLFFPYSPLSFVLTTFMNF
KLFLNLQFWSWFFVNSKILKFTLFMNFSLGGLPPFLGFLPKWLVIIQQLTLCNQYFMT
IMMSTLITLFFYLRLICYSAFMMNYFENNWMKMMNSINYNMIMTFPSIFGLFLI
SLFYFMP"

tRNA
1264. .1329
/gene="mt:ND6"
/product="tRNA-Trp"
/db_xref="FlyBase:FBgn0013685"

tRNA
complement(1322. .1383)
/product="tRNA-Cys"

tRNA
complement(1403. .1468)
/product="tRNA-Tyr"

CDS
join(1470. .1472,1474. .3009)
/codon_start=1
/exception="mechanism underlying reading frame shift after first codon uncertain"
/transl_table=5
/product="cytochrome c oxidase subunit I"
/protein_id="AAC47812.2"
/db_xref="GI:7412849"
/translation="MSRQWLSTNHKDIGTYFIFGAWAGMVGTSLILIRAEILGHG
ALIGDDQIYNVITTAHAFTIMIFFVMPTMIGGFGNWLVPMLGAPDMAFPRMNMSEW
LLPALSLLLVSMSVENGAGTGWTVYPLSAGIAGGASVDLAIPLSLHLAGISSILGA
VNFITTVINMRSTGSLDRMPLFWVSVITALLLLSLPVLGATIMLLTDRNLNTSF
FDPAGGDDPILYQHLFWFPGHPEVYIILPGFGMISHIISQESGKRTFSGLMIIAM
LAIGLLGPIVWAHMFVGMVDVTRAYFTSATMIIAVPTGIKIFSWLATLHGTQLSYS
PAILWALGFVFLFTVGGLTGVVLANSVDIILHDTYTVVAHFHYVLSMGAVFAIMAGF
IHWYPLPTGLTLNNKWLKSHPIIMFIGVNLTFPQHLGLAGMPRRYSYDPDAYTTWN
IVSTIGSTISLLGILFFFFIIEWSLVSQRQVIYIQLNSSIIEWYQNTPPAEHSYSELP
LLTN"

tRNA
3012. .3077
/gene="mt:ND6"
/product="tRNA-Leu"
/db_xref="FlyBase:FBgn0013685"

CDS
3083. .3767
/note="TAA stop codon is completed by the addition of 3' A residues to the mRNA"
/codon_start=1
/transl_except="(pos:3767,aa:TERM)"
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="GI:1166532"
/translation="MSTWANLGLQDSASPLMEQLIFFHDHALLILVMTVLVGYLMFM
LFFNNYVNRFLHGLIEMIWILPAIILLFIALPSRLLYLLDEINEPSVTLSIGH
QWYWSYSEYDFNNIEFDSYMIPTNELMTDGFRLLDVDRVVLPMNSQIRILVTADVI
HSWTVPALGVKVDGTPGRNLQNTNFINRPLGYQCSEICAGNHSFMPVIESVPVNY
FIKWISSNNS"

tRNA
3768. .3838
/gene="mt:ND6"

FEATURES	Location/Qualifiers
source	1. .4601 /organism="Drosophila melanogaster" /organelle="mitochondrion" /strain="Oregon-R" /db_xref="taxon:7227" /dev_stage="embryo"
gene	1. .4601 /gene="mt:ori" /note="mitochondrial origin" /allele="" /db_xref="FlyBase:FBgn0013687"
repeat_unit	650. .1022 /gene="mt:ori" /note="repeat I-A" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	1023. .1360 /gene="mt:ori" /note="repeat I-B1" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	1361. .1705 /gene="mt:ori" /note="repeat I-C/A" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	1706. .2043 /gene="mt:ori" /note="repeat I-B2" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	2044. .2388 /gene="mt:ori" /note="repeat I-C" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
misc_feature	2491. .2511 /gene="mt:ori" /note="deoxythymidylate stretch" /db_xref="FlyBase:FBgn0013687"
repeat_unit	2512. .2648 /partial /gene="mt:ori" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	2649. .3112 /gene="mt:ori" /note="repeat II-A" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	3113. .3576 /gene="mt:ori" /note="repeat II-B1" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	3577. .4040 /gene="mt:ori" /note="repeat II-B2" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem
repeat_unit	4041. .4504 /gene="mt:ori" /note="repeat II-C" /db_xref="FlyBase:FBgn0013687" /rpt_type=tandem

Qy 14 aatttttaataagtagaaacnctaaccaattttttaataataaagctgactcctagtacaag 73
||| ||||| || | | | | | | | | | | | |
Db 1259 AATAATTAATAAAAAATTTTTTAAATATAAATAAAATTAAAATGATTTTATATAAA--- 1315

Qy 74 cttttatctattcttctattttgtcttcctctaggcttgccaatcgagaattttcttgt 133
|| ||||| || | | | | | | | | | | | |
Db 1316 ATTCAAATTCATATATTATATATATATACATATAATTTAATTTCAATTAAATTATATAA 1375

Qy 134 ttacaataataataacatcgtagaanaatttttatcaaattgaagtottaaccatc 193
||| ||||| || | | | | | | | | | | | |
Db 1376 GTATAATAAAATAAATTTATTTTAATCTACTAAATCTGAATTAATTGTTATATATATAT 1435

Qy 194 tttaatattgttagatgtaatttaaagaagaataacatatctctggacatgtatt 253
|| | |||| | | | | | | | | | | | |
Db 1436 ATATATATATATGTAAAAAGAAAATAAATTTATCCCCCATTCATAAATTTATTATA 1495

Qy 254 tcactctaattgt---ttgtggcttgggtgatagggtattgatgtacgatgtctttaa 309
|| | || | | | | | | | | | | | | | | |
Db 1496 TAATTAACAATTAATAAAATATTTTTTAAAAAAAAAATTATTATTAAATTATACTTA 1555

Qy 310 atcacatatcacattttgagtttgatgatgataagtcgcataancgaaaattggtgtg 369
|| | || | | | | | | | | | | | | | | |
Db 1556 ATAAACTATTTTTATAATAAATTTATTTTATAAAATAAATTTATTAATAATTAATAAGA 1615

Qy 370 atcttcacttttgaaccttgatgaagccaccaactttaacaagtttgattgtgcatata 429
|| | || | | | | | | | | | | | | | | |
Db 1616 AATATTTTTATATAATAAAAAATTAATAAATTTTAAAAAATCAATTTATATTATATA 1675

Qy 430 tatatatatatcttcaatttttataaaaaaattgtgtttaaataatttcacagtattatt 489
||||| ||||| || | | | | | | | | | | |
Db 1676 TATATATATATATATATAATTTTAATTTCAATTAAATTTATATAATAATAAAATAAT 1735

Qy 490 attttttatctctaatttttattgtgccaaatttttagtgatattttaacataaaaa 549
| |||| | |||| | | | | | | | | | | | |
Db 1736 TTATTTTAATCTACTAAATCGAAATAATTAATTATATATATATATATATAAAAAA 1795

Qy 550 aaattgtcacacatttacagcccataacaaataattatataaatttcattaaaaaata 609
|| | || | | || | | | | | | | | | | | |
Db 1796 TGAAAAATAAATTTATCCCCCATTCATAAATTTATTGTATAATTAAACTTAAAAAATA 1855

Qy 610 tattttaaatataggatataaataaactattttagaattattctacttttaagataacata 669
|| | || | | || | | | | | | | | | | | |
Db 1856 TTTTTTTTTTAAAAAAAATGATTATTAAATTAATACTTAATAAACATTTTATATAATA 1915

Qy 670 gggttaagtataaattaagaaggttagtttttgaagatgagtatatatgtcgtgaac 729
| | || | || | || | || | || | || | || | |
Db 1916 ATTATTTTATAAATAAAATTTATTAAATAATTAAATAAAATTTTAAATAATAAAAA 1975

Qy 730 ataactactaacattttttataactcttctggttttgaagttccaaaaagaaaatgga 789
|| | || | || | || | || | || | || | || | |
Db 1976 ATTTAAAAATGATTTTTTATAAAAAATCAATTCAATATATTATATATATATAC-ATATAAT 2034

Qy 790 ggaattgtgagagtaagttcatgtttatattatcacataatgaagttgatgtttctctt 849
|| | || | || | || | || | || | || | || | |
Db 2035 TTAATTTTCAATTAAATTATATAGTATAATAAAAAATTTATTTTAACTAAATCTG 2094

Qy 850 tttaatatttttatacaaaatttttaataaaaaataaaggattgaatgaaaaataaa 909
|| | || | || | || | || | || | || | || | |
Db 2095 AATTAATTAAATGTATATATATATATATATATATATATATGTAATAAGAAAATAAATTTAT 2154

Qy 910 tgaaagtcgtttactaataagtcattatgcatgttttgcgcactacttaataatagata 969
|| | || | || | || | || | || | || | || | |
Db 2155 CCCCCTATTCAATAAATTTATATATAAATAAACTAAAAAATTTTTTTTTTAAAAA 2214

```

misc_feature      complement(39996. .40005)
                  /gene="garp"
                  /note="potential splice acceptor sequence"
misc_feature      complement(40204. .40209)
                  /gene="garp"
                  /note="potential splice donor sequence, aag/gtaaca"
gene              45401. .50233
                  /gene="MALIP3.07"

ery Match          5.8%;   Score 175.8;   DB 33;   Length 67970;
st Local Similarity 43.9%;   Pred. No. 7e-11;
tches 1220;   Conservative 1;   Mismatches 1539;   Indels 22;   Gaps 10;

139 atataataaatcatcgtagaataaattttattcacaattgaagctctaaccattctttaa 198
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
7824 ATAAATAATATTCCCTTAATTATATTAACATTATTAAAAATATAAATTAAGGAATAAATATT 7883

199 tattgtagatgtaatttaaatgaaagataaatacatattcttgacatgtattttcatc 258
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
7884 TTCTTAAAAAACGATGATATAATAAATAATTTAATATATATTAAATTCATTATATATATA 7943

259 ttaatgtttgtggctttgtgatagggtattgatgtacgatgtcttttaaacacatat 318
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
7944 ATAATATATTTACATTAAATTTTAATAGTTTATATATA---ATTATATAATTTCTTTAT 7999

319 cacattttgagtttgtatgatgataagtcgacataancgaaatattggtgtgatcttcact 378
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
8000 TAATAATTAATTATAGTACACATTTATATTAATAATATAATTAAAGAATATTTATCTT 8059

379 tttgaactttgataagtcaccaaaacttttaacaaagtttgattgtgtacatatatatat 438
   || | | | | | | | | | | | | | | | | | | | | | | | | | | |
8060 TTGTATTATATATATATATATTTAAATGGAATAAATTAATAATAAATAAATAAATAAATAA 8119

439 atcttcaaattttataataaaaaattgtgtttaaataatttacagttatatttttttta 498
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
8179 TATATTAAATATATAAATTAAACAAAATAATATAATAATTAATAAATAAATAAATAATATAT 8179

499 tctctaatttttttgtgcgcaaatttttgattgtatatttaacataaaaaaaaaattgtac 558
   || | | | | | | | | | | | | | | | | | | | | | | | | | | |
8180 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8239

559 acatttacaagcccatatacaataaattatataaaattattcattaaaaaatatatttaaat 618
   || | | | | | | | | | | | | | | | | | | | | | | | | | | |
8240 TAAAAATATTAATAATATATAAATTAAATAATATAAATAAATAAATAAATAAATAAATAAATA 8299

619 ataggatataaatataactatttttagaattattctactttaagataacataggtttaaag 678
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
8300 AATTAAATAAAATATTAAAAATATATAAATTAAATAATATAAATAAATAAATAAATAAATAA 8359

679 tataattaataaggttagtttattgttaagatgagtatatatgtcgtaaacataatcact 738
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
8360 TATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8419

739 aaccatttttattaacttcttggttttggaagttccaaaaagaaaatggaagggaatttg 798
   || | | | | | | | | | | | | | | | | | | | | | | | | | | |
8420 AAAATATATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8477

799 agagtaagttcatgttttatattatacataaatgaagttgatgttttctcttttaattatt 858
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
8478 ATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8537

859 tttatacaaat-atttaataaaaaataattaaggattgaatgaaaaataaatgaaagtc 917
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
8538 ATTATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8597

918 gttttactaatagtcattgttcatttggcgcactctacttaaaataagataaaattaatt 977
   || | | | | | | | | | | | | | | | | | | | | | | | | | | |
8598 AATAAATAAATAATATATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8657

978 gtggtacattagatcaaaagaacaaactagatttggccattcttattgttaaaagctggt 1037
   || | | | | | | | | | | | | | | | | | | | | | | | | | | |
8658 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8717

```

Db 41588 AATTTTTTTTTTAAATATATATATATGTTAATTTATCTAACCCATTAATTT 41529
Qy 182 gtcttaaccatcttaaatattgttagatgaatttaaatgaaagataaacatattctt 241
Db 41528 TTTTAAAAATATTATATTATTAATCTTTTAAATACTGTATAAAATAAATGAATTTTT 41469
Qy 242 ggacatgtattttcatcttaagtgttggtggt---gatagggtattgatgtacg 298
Db 41468 TTCAATTTAAAAATTTATATATGTAATGTATAAAGAAATATATATAAATTAAT 41409
Qy 299 atgtcttttaaatcacatcacattttgagttgtatgatgaagtcacataancga 358
Db 41408 ATATATTTGTATAATAATAATATATTAACCTAATAAACAATAACATATAATAA 41349
Qy 359 aataggtgtgatcttcacttttgaaacttgataagtcaccaactttaacaaagttga 418
Db 41348 TTAATTTAGCAATTTATTAATTAATAATATATTAATTT-CAAAAATTTATATATAATA 41290
Qy 419 ttgtgtacatatatatatatcttcaaaattttataaaaaattgtgttaataattt 478
Db 41289 ATAAATACACATCAATA-AAATAATCATTTTATAAAATAAATAATAAATAAATTTAAT 41231
Qy 479 acagttatattttttttatcttaattttattgtgcgcaaaatttttagttgatatt 538
Db 41230 TTA--TATAATATAATAATATTATAAAATACAATTAATAAATTTTCAATAAATTT 41173
Qy 539 taacataaaaaaattgtacacatttacaagccatatacaaaataattatataaatttc 598
Db 41172 AAATATACATATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 41113
Qy 599 attaaaaatatatttaaatataggatat--aaataaactattttagaattttctactt 657
Db 41112 TATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 41053
Qy 658 taagataacataggttaattgtataaataaaggttagttttattgaaagatgagata 717
Db 41052 ATTAAATACATATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40993
Qy 718 tatgtcgtaaacataactactaaccatttttataacttcttggtttgaagtccaaa 777
Db 40992 ATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40933
Qy 778 agaaaattggaagggaatttgagagtaagttcatgtttatattatacataatgaattga 837
Db 40932 ATTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40873
Qy 838 tgttttcttcttttaattttttatacaaaatatttaaaaaataaaggattgaa 897
Db 40872 AATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40818
Qy 898 tgaaaaaataatgaaagtcgttttactaatagtcataattgcatttgtcgcacttact 957
Db 40817 TTAATTATATTATTAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATA 40758
Qy 958 aaataatagataaatttaattggttacattagatcaagaacaaactagattttgccc 1017
Db 40757 TAATTATATTATTAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATA 40698
Qy 1018 ttctattgtttaaagtcgttcgtttacattaaataaggtacattgtacatgcccgta 1077
Db 40697 TTATTTAATAAATAAATAAATAAATGT--AATAAATATATATAAATAAATAAATA 40640
Qy 1078 taactatctggttattctatcaatcacgtaatttttaacagtagaataatgaattt 1137
Db 40639 TATTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40580
Qy 1138 ttaaatagaaagggtcaaattgtttattgatctaacacgtaggattaatttacttatt 1197
Db 40579 TTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40520
Qy 1198 tcctaagaaataagtaaaatataatttgaatcttaatacaaaaactttcatgatacttt 1257
Db 40519 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40460

Qy 1258 tatcatattttacttataatttaattgtgagagtaacaaartaaaaacatagaaac 1317
Db 40459 TTTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40400
Qy 1318 accaaaagtagttatggtgtgactcatatacacagttaaaaattgaataaattttttc 1377
Db 40399 AAGTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40340
Qy 1378 ttctgcatttaattccatcatgggtttttttttctagtttaagccataattatcaa-at 1436
Db 40339 TTATAATATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40280
Qy 1437 aatcatcatttaattcctcaataccccgcctgctccctccctcaacttaaacccaa 1496
Db 40279 TATAATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40220
Qy 1497 ctaaccccagcaccaacgcacttttaagccactattttctagccatgctcctgcaact 1556
Db 40219 TTAATAATACGTGTGTCTAATATATATATATTATTAAATTTAATTAATTTGATTCAT 40160
Qy 1557 taaagaaaagtaagctaacctgcaatcattccat-atcgaggccctcaacagataaagtt 1615
Db 40159 AATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40100
Qy 1616 ggttgatgggtttgcaccaagttgttaaaacccgcccctcaacttccctttcttttcat 1675
Db 40099 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 40040
Qy 1676 cctccccactccacacccctcaattttctcatatggtttctattataagttctttataat 1735
Db 40039 TATTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39983
Qy 1736 cacagaatcaagataagtcctcagcaacaaaaaacctagctctcagcaagatctgga 1795
Db 39982 CATTTAATTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 39923
Qy 1796 ctagtccagagctctgaatattggtacattattacagtcacaaacagttacaaagctgt 1855
Db 39922 TGAATTAAA-ATGTTATTATAATATTATAAATAAATAAATAAATAAATAAATAAATA 39864
Qy 1856 tgcagataaacactgaatctgctatagttgtttttggtttacatagttccacgtgaaa 1915
Db 39863 TTAATAATTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 39804
Qy 1916 ctatgaagcatctctaagaaaacccaaactatcatcaacccatcgatcaatgaatcga 1975
Db 39803 AAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 39744
Qy 1976 tttcaatttttcagctataagtttcttttttaacttttttttacttattttatacga 2035
Db 39743 ATTAATATAAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 39684
Qy 2036 attctatggataatgttccctacaacatgtcattacaagtgttaattataaattccatt 2095
Db 39683 AAAACATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 39624
Qy 2096 cttctattttactaagattagtaacttcaactgctgatttttactaattttatttt 2155
Db 39623 TTGAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 39564
Qy 2156 ataaattgttagaatgattttttcaataatttaacaaatatttaattattattatta 2215
Db 39563 TTAATTATTATA----TTTATTATTATTATTATTATTATTATTATTATTATTATTATT 39508
Qy 2216 ttattatttctcaatttttttaaaacaaaaacataaatttttgacaaataaaaaaattg 2275
Db 39507 TAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 39448
Qy 2276 aattaatttctcaattttttgcaactattacaaaaacccctcatagtcctaatcttaa 2335
Db 39447 AAATAATATATTAATTTATTATTATTATTATTATTATTATTATTATTATTATTATT 39388

Db 9105 TGTTTTATTATTAAATTATTATATTAATATATTTTTATTATTAAAAATTATATATTA 9046

Qy 1712 gttctattataagttcttttataatcagagaatcaagataagctcctcagcaacaaaaaac 1771

Db 9045 ATAATAAAGACAATATATTAAACATATAAATTAATATAATTTTTATTATTATTAAT 8986

Qy 1772 catggctctcgcagcaagatctcgactagtcagagctctgaatttggatcattattacag 1831

Db 8985 TTGTTTAATATATTATTATTTTATTAAATTATTATTATAATATTATTTTATTATTA 8926

Qy 1832 tcaaaaaacagttaacaaaagctgttcgcagataaacactgaatctgctatagtttgtttt 1891

Db 8925 TTATTTAATTATATTATTATTTTATTATTAATTATATATTTTAAATATATTATTTATT 8886

Qy 1892 gggttt-----acatagtgtccacgtgaacctatgaagcatctctaaagaaacccaaact 1945

Db 8865 TATTTAATATAATATTATTTTATTTAATATAATAATTTTAATTATTTAATATAATAT 8806

Qy 1946 atcatatcaaccccatgcgatcaatgaatcgatttcaattttcgcagtataagttc-----c 2000

Db 8805 TTCATTTAATTCATTTAATATAAATTTTAAATTATATTATTAATTATATATTTTAAACAT 8745

Qy 2001 ttttaactccttctttttacttcattttataacgaattctatggataatgttccctacaa 2060

Db 8745 TTTATTTAATTATTTAATATATAAATTTTATTATTATTAATTATATTATTAATTATA 8686

Qy 2061 acatgtcattacaatgttttaattataaattccattcttctattttactaagattattagta 2120

Db 8685 TATTTAATATTTTATTTAATTATTTAATATAAATTTTATTATTATTAATTATATTATA 8626

Qy 2121 acttcaaacctgcgtatttttactaatttattttataaaattgttagaatgattatttt 2180

Db 8625 ATTTATTTAATATAATATTTTATTATTATTAATTATATATTTTAAATTATATTATTTA 8566

Qy 2181 caataaatttaacaacaatatttaataattattattattattttctcaattttttattaaa 2240

Db 8565 ATATTTTATTTAATTTTATTTAATATAAATTTTATTATTATTAATTATATTATTAATT 8506

Qy 2241 caaaaacataaaatttttgacaaattaaaaaataatgaattatttctcaatttttcgtgca 2300

Db 8505 TAAATATAAGTATTATTTTATTAAATATATATATTTATTAATTATATATTATTAATTTT 8445

Qy 2301 actattacaaaatccttcatagtctcaatcttaatttgatgcagaggatgataataatct 2360

Db 8445 ATTTAATTATTTAATATAATATTTTATTATTATTAATTATATTATTAATTATATTTTA 8386

Qy 2361 taatttgatgcagaggtaataatggcgcggtttgagctggactta---agcatgat 2416

Db 8385 ATATTTTATTTAATTTATTTAATATAATATAATTTTATTATTAAATTATATTATTAATT 8326

Qy 2417 tgacgtacttttatatttttccaaattcaaccagctcgaaatagagctctaaaaatttgt 2476

Db 8325 TATATATTTTAATTTTATTATTTAATTATTAATATAAATTTTATTATTATTAATTATATT 8266

Qy 2477 ccaatttaatccaagccatttttaagttcgtccatatttttttaatttaaaaaattta 2536

Db 8265 ATTAATTATATATTTTAAATATTTTATTATTAATTATATATTATTAATTATATTTTATA 8206

Qy 2537 tatcattttatttttaatttttaattttttatatattttttattttattgaaaaattttat 2596

Db 8205 ATATTTTATTTTTATTTTTATTATTTAATATAAATTTTATTATTATTAATTATATTAT 8146

Qy 2597 atagtcactcttaacattatgttaatgtttatattagagtagtattatatatttttagtat 2656

Db 8145 TTGCTTAATTTATATATTTTAATATATTTTATTATTATTTTATTATTATTAATTATTC 8086

Qy 2657 aggtttatttttgtaataaaacttaaaatgggtctgttggygctagactggaccttaaat 2716

Db 8085 A--TTAATATATATATATAATAACAAAGATAAAATATTCCTTAAATTATATTATTA 8028

Qy 2717 gctcaaacctcaaaacttaattcatattttaaacaggcttaaatttttattttacactgttt 2776

Db 8027 TATAAATGTGACTATAATTAATTTATTAATAAAGAAATTAATAATTTATATATAAATAT 7968

Qy 2777 caaattttctgggtgaaatatcttcgagtgtagattaataacaccacaggctctaattga 2836
 | | || | | | | | | | | | | |
Db 7967 TAAATTAATGTGAAATATATTATTATATATAAATGAATTAATATATTTAAATATT 7908
 | | | |

Qy 2837 tgctcaatga 2846
 | | |
Db 7907 TATTATATCA 7898

```

RESULT 17
PFMAL3P5/c
LOCUS PFMAL3P5 86829 bp DNA INV 11-FEB-2000
DEFINITION Plasmodium falciparum MAL3P5, complete sequence.
ACCESSION AL0034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162
AL010206 AL010210 AL139179
VERSION AL034556.2 GI:4493931
KEYWORDS HTG.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 86829)
AUTHORS Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T.,
Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,
Hornsby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J.,
Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,
Rajandream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S.,
Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and
Barrell,B.G.
TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum
JOURNAL Nature 400 (6744), 532-538 (1999)
MEDLINE 99376085
REFERENCE 2 (bases 1 to 86829)
AUTHORS Bowman,S., Skelton,J., Churcher,C., Lawson,D., Quail,M. and
Barrell,B.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 86829)
AUTHORS Lawson,D., Bowman,S. and Barrell,B.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Mar 24, 1999 this sequence version replaced gi:4034877.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/Projects/P\_falciparum.
FEATURES
source
1..86829
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="3"
/clone="MAL3P5"
gene
324..2944
/gene="MAL3P5.1"
CDS
join(324..668,1199..1303,1460..2944)
/gene="MAL3P5.1"
/note="predicted using hexExon; MAL3P5.1 (PFC0575w),
Hypothetical protein, len: 645 aa"
/codon_start=1
/protein_id="CAB38969.1"
/db_xref="GI:4493933"
/db_xref="SPTREMBL:O97258"
/translation="MYLKNVYIYSSCFILFDLCFSFHLKMKYKNHNMNMKSVTFFL
RSPQIYKRKRFRSRIKNVFSFKKQKKPLFLFENLKKGSFSLGFRNQYDQVKNREKKK
KKKKKKKKKNPKVHSILNQISEKVKKKAENYLAHLFLFKDENITLFSMMHIMDF
FASKQKVICERIDIRSKKKRKNLSIYINLPICITLIYFTYCMCLLIKYISHLCIFPPP
FFCFILCYNLIERIYVECVGDLIRKKIERYNLYCEKKKKIFPHMKDAIKKMEINMKDD
LYFNHYDELRLCFTMKLNIERNKNKIIRSNYDNIWNDISIDKDMYMNPNIDVNIWNI
SLDEKIKQFENPDDENLKLKDTYEQQLFNQNIKYIEEDQPLYNINDNSININDNN
NNINTMKNKHKIKDTYNDDDYDEKEEDLVIQKNIDDIYKNTIGMKNLSLEEFKNQF
IPOADIEPONFLSNVNDLDOGRVKSNDENTKSTEHIKNKNINKGYDTLIONOMENI

```

TITLE Direct Submission
JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT On Dec 16, 1999 this sequence version replaced gi:5763807. For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

FEATURES
source Location/Qualifiers
1. .67970
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="1"
gene complement(1748..3276)
/gene="MAL1P3.01"
CDS complement(join(1748..2598,2748..2848,2990..3276))
/gene="MAL1P3.01"
/note="MAL1P3.01, conserved hypothetical protein, len: 412 aa, similarity: UPF0006 family eg to YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa overlap)"
/codon_start=1
/product="conserved hypothetical protein, UPF0006 family"
/protein_id="CAB63556.1"
/db_xref="GI:6594244"
/translation="MKLVFHYIKYINVLFYISIIPLKSNLSKIYNDLRYISTVKNKYV LQIKKRSNLAKNHNIRKMDNESSFDIGSNLTKMFDGVNSKKHENDLQVNLRAK NNVVDKIIITCTCLAEIDKSLKICETVDPEGKFLYL SAGVHTNCEYFIDKKNKHEEKE IIAKKEYEPIYKFNQEVNSKMGNGKKICDGEKDMNMLELLEKNLDTIPGPKY NEKDKEYLENLKNKIIKYNNRIVCIGEIGLDFDRLYPCSKYIQYVIFQLKLVQMFN LPMFLHMRNCSETFFKIVDIYKFLFEGKNGGVIHSFTDKEDIVHIIYQYKNLYIGVNG CSLKSLNINAVKKIPLNLLLETDAPWCGVKKTHASVEYIKDYEKRAYTNLKKIKN IIKCDDNTIFKERNPEYNA"
misc_feature complement(2599..2610)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"
misc_feature complement(2742..2747)
/gene="MAL1P3.01"
/note="potential splice donor sequence, atg/gtataa"
misc_feature complement(2849..2861)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"
misc_feature complement(2984..2989)
/gene="MAL1P3.01"
/note="potential splice donor sequence, aaa/gtaaaa"
gene 5005..5496
/gene="MAL1P3.02"
CDS 5005..5496
/gene="MAL1P3.02"
/note="MAL1P3.02, hypothetical protein, len: 163 aa, contains possible signal sequence"
/codon_start=1
/product="hypothetical protein, MAL1P3.02"
/protein_id="CAB63557.1"
/db_xref="GI:6594245"
/translation="MKLLNRRFVLCPIIILFFFLNSVVLGNNRRNNINPHETENAAK AMRKLSSGEINSIKLDNGDELKIKLNDKHKDSTKWDKSYFISNLEEEKYSQTDLFR KKQEIENANTKIIEDRQEFYILNNDENIATRFVLENNFDELYIQSFQKSLIDIQS LNN"
misc_feature 8020..10389
/note="possible cen1, region of very high [A+T] content"
gene 14884..20352
/gene="MAL1P3.03"
CDS 14884..20352
/gene="MAL1P3.03"
/note="MAL1P3.03, putative ABC transporter, len: 1822 aa"

/codon_start=1
/product="putative ABC transporter"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/translation="MTTYKENVGISNKGNNKKKSCQNISFLNLSFDWIRPLINDLIK GDIQELPNICRNPDPYVYASKLEENLRDIEVEDSEFYSEKSSNEHVLHCNSNDASE KKVYNVYHNILWSILTKFKFRILLIISFYILETLIVTLGGKFIDYMRILEGQKIPV YISFLKDFKVFSGLVVVMIMFHLFFFEALLHFFHLFTINLKVSLMYFLYKINLCSNN NHLQNPDAFYNTYRKFSQSTEIDRDLFSIGKNASSSSSGIKNNKNNDNNKFEVEN DYIINFIKSTKMEKDSLNNRSLPNVNIYIMFSDVPSVTFVTSCINLNFNVFKIF MSFYVFIKIGSNGVIAIWSIALYSAMILFEFLPSLFSKYLIYRDKRIDNMHVL KEFKLIKMFNWFSAFKYINIFRMKEMKYCKIRLYLSNIGVFISISSSDIVEVVFIF IYKDRNLKKEEIKFTSIIMPLVYKILISVANFPLNVNMEGIVNKRNNYINDH LYNDIKNYFMYRTYRNYEDYNIIVDKTLQENITSHDDGTSNHLKHLKNVKNKLTN MFKYFFFYHKMNYHKNIIINQILSGLLKNVDDNTNKKICQEHKSNSTYNNSSHIHE KKEEYENIHNSNSTMNEPEKKEKKNNNEYIKLENCSPGLSDNKCNDHILKNINFN LKRNSLAIIGNVSGSKSAFFHSILGDPNTHGNLYIENFFKKMPILYVPQNSWLFMG NRSMLIFGNEYNPLIYKYTILOSELNLDSTIEGDMKYINDHNLKSGQKVRICLA RALYEHYIMHKLCTDYEKLLIQNEILDKDLINNNKISSYNNKSKLVNYPNFEN YLQKCLMDDNNFYLLDDIPTSLSIKKIFSLCKEDNISFKDNCSPFIISMNKS TLDNFLIEDLDNVQYEVNIFDIQDKTLKYRGNISYEMKNLNLITKESHWGYSNLT IDYTRIKLDFEVLNHHVKSNNKMYKEAYFVGKNTESVSFEIDSINKYEIKMKKKNY KKEHMKNNKDNNNNNNNNDKHININMNDNERNYNDINLGNPTDSDPTVSSLGNE YTLDTYTSNNSDKEEIVKPLYKDTHEEFNKSSSMFPVKSSSMNNPNFNFYEDNSSS FKGSISLETYLVWFGVQGVFLVLTYSVIFMLISITFYDEIKFVFLTMSSIIKNNKESD TILQKQVRYLEYFVILPIISLVTSGICFSMIYGNITSIAKVHNNILYSILNAPLYIF YNNNLGNIINRFIIDSAPDYGFLKRIYKAFFIFRCILSSLLIYIMRDCIFIFPFV IILYIFVFKRFSRGCKEAQRLYLSCHTPLCNISNALSGKNININIKNTYHLDVYE HYINNFRISYFFKWLINIWASLTYKIVFILLITTYIIMPHLYASGLIKLYKEKNTVRI LSTLGVCISFSARLGVIIKFLCDYTHIEKEMCCVQRLKEEFKISNKENASMNKENEL NVITTQTYKEKNENISDKISAIVEYKYNVSSLSIINSQDDSEKKYKIGFENVYVSFK KKIPLVNGTYKIDEEPSLKNINMYALKNQKIGIVGKSGAGKSTILLSILGLINISQ KITVEGRDIRTYNRKGEDSIIIGLAQSSSFVFNWNIPTIDPYNFTDDEIVHALKLN GINLGKNDLYKMHKQDMKSNYKIIQTSKVINQSDNTILLTNDICIRYLSLVRVLYN RKYKIIILDEIPFNWLNWSVHDELNSFLIGAKASFNYIIRNHPNNTVLIISHHANT LSCDDYIYVLKGEITYRCSYEDVKQTSLSHLEMD"

rRNA 23896..31533
/gene="rRNA"
/note="region containing small subunit, 5.8S and large subunit rRNA genes and spacer regions"

gene 23896..31533
/gene="rRNA"

gene complement(31966..32775)
/gene="MAL1P3.04"

CDS complement(join(31966..32476,32675..32775))
/gene="MAL1P3.04"
/note="MAL1P3.04, conserved hypothetical membrane protein, len: 203 aa, similarity: P. falciparum chromosome 2, PFB0110W, O96126 predicted integral membrane protein (255 aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity in 191 aa overlap)"
/codon_start=1
/product="conserved hypothetical membrane protein, MAL1P3.04"
/protein_id="CAB63559.1"
/db_xref="GI:6594247"
/translation="MKKSYTFINVTILLFLTLLEFLTYNNYDTFSKTKFNNKIDIN RPKRIIAEASEEQKYPWEEDFLCILNEELRPEHNDSPYLPHEYENIDKINLSINS TKIKWETIKKMRQNYEKETDNMNNWRDFMWHYKWNANIYLYKVHKLINITLKDLTNPI HDKEETITTWIKWQIEDIEYFLNQLQVWLRILTELFYKNKE"

misc_feature complement(32477..32486)
/gene="MAL1P3.04"
/note="potential splice acceptor sequence"

misc_feature complement(32669..32674)
/gene="MAL1P3.04"
/note="potential splice donor sequence, aaa/gtatat"

gene 36657..37343
/gene="MAL1P3.05"

CDS join(36657..36743,36864..37343)
/gene="MAL1P3.05"
/note="MAL1P3.05, hypothetical protein, len: 188 aa"
/codon_start=1
/product="hypothetical protein, MAL1P3.05"

Qy 1205 gaataagtaaaatataatttgaactcttaatacaaaaaaccttcgatcatttttatcata 1264
| | | | | | | | | | | | | | | | | |
Db 7950 ATTATTA-TTATATATATATAATTATTATTATTTTAAATGTTAAATAAATAAAAAATA 8008

Qy 1265 ttttacttataattttaatttgtagagtagaacaaaattaaacaacatagaacacacaaaa 1324
| | | | | | | | | | | | | | | | | |
Db 8009 TAATAAAATTAATTAAATTAATAAAAATAAATAATTATAATTATAATATTACATATA 8068

Qy 1325 gttggttatgggtgtagctcatatcacagtgtaaiaatttgaataaattttttcttcgtca 1384
| | | | | | | | | | | | | | | | | |
Db 8069 AAATATATATATTATATATATTTAATTATTATTATTTTATTTAATGAATAATAATAAT 8128

Qy 1385 ttaattccatcatggggttttttttttctagttaaggcataattatcaaataatcatca 1444
| | | | | | | | | | | | | | | | | |
Db 8129 TAATTATATTAAATTAAATTATATTAATAATATTATTATATATATTAAATTAAATTAATA 8188

Qy 1445 ttaactctatcaataccccgcctgacctcccctccaacttaacttaaccaccaactaacacc 1504
| | | | | | | | | | | | | | | | | |
Db 8189 TATTTATTTTTTAATA----AACTTAATTAATTAAATTATTATTATATAAAAAAATATATT 8243

Qy 1505 cagcaccaaacgcactttaatagccacactattctagccatgtccttgcacttaagaaga 1564
| | | | | | | | | | | | | | | | | |
Db 8244 AAATTATGATATATATATAATATTAATATATATTAAATTAATTAATTAATAAAAAATA 8303

Qy 1565 agtaagctaaccctgcaatcattccatatcgaggccctcaacagataaagggttgtagtg 1624
| | | | | | | | | | | | | | | | | |
Db 8304 TATTAAATTATTTTTTAATTAATTTATATATATTATTAGAGTCATTATTATTATTATT 8363

Qy 1625 gtgtgcaccaagtggttaaaacccggccctcaactccctttctttcatcctccccac 1684
| | | | | | | | | | | | | | | | | |
Db 8364 TTTTAAATTAATTAATTAATTTATATATTTAATAATATATTTTGTTCTTTGTATTAAAT 8423

Qy 1685 tccacacccctccaattttcttcctataggtctctatt-----ataagttctttataa 1734
| | | | | | | | | | | | | | | | | |
Db 8424 ATATTATTATTTTATTATATTATTAATTAATTAATTAATTAATTAATAAAAAATTATAATA 8483

Qy 1735 tcacagaaatcaagataagctcctcagcaacaaaaaacctggctctcgagcaagatctgg 1794
| | | | | | | | | | | | | | | | | |
Db 8484 TTATTTCAATTATTTTAAATATATATTATATAATAATAATTATTCCAAATATAATAAAT 8543

Qy 1795 actagtcagagctctcgaaatttggtacattattacagtcaaaaacagtgtaacaaaagctg 1854
| | | | | | | | | | | | | | | | | |
Db 8544 AATGTATTATTAAAGTTAATAATAATTATTAATTTATTTTATAAAATAAATATTGCATT 8603

Qy 1855 ttgcagataaaacactggaatctgctatagtttgtttttgttttacatagtgtccacgtgaa 1914
| | | | | | | | | | | | | | | | | |
Db 8604 AATTCATTAAATATAATTTATTATATA--TTATTTTAAATAAATTATGATATAATATTA 8661

Qy 1915 acctgaagcatctctaaagaaacccaactatcatatcaaccctcgatcaaatgaatcg 1974
| | | | | | | | | | | | | | | | | |
Db 8662 TATATATTAATAATTTAATTAGTATATAAAAATTTAAACATATACATTATAATTACACA 8721

Qy 1975 atttcaatttttcgcagataaagttccctttaactctttctttttacttcattttataacg 2034
| | | | | | | | | | | | | | | | | |
Db 8722 TAATTATTATTATTATTTAATTAAATTTTACTATTTTTAAATTAAATTATTATTATTAT 8781

Qy 2035 aattctatggataatggtccctacaaacatgtcattacaagtgttaattataaaattccat 2094
| | | | | | | | | | | | | | | | | |
Db 8782 ATTATTATTATTATTATTTAATTAAATTATTAAATATATAATTATTATTATTATTATT 8841

Qy 2095 tcttctatttttactaagatatttagtaacttcaaaactgctgatttttactaattttatt 2154
| | | | | | | | | | | | | | | | | |
Db 8842 ATT 8901

Qy 2155 tataaattgttagaagattatttttcaataatttaacacaatatttaattattattatt 2214
| | | | | | | | | | | | | | | | | |
Db 8902 ATTAATTATTTAATATATAATTTATTTTATTATTATTATTATTATTATTATTATTATT 8961

Qy 2215 attattatttctc-aatttttttaaaacaaaaacataaatttttgacaaattaaaaataa 2273
| | | | | | | | | | | | | | | | | |
Db 8962 ATTTTATTATTATTATTATTATTATTATTAAATTTATTATTATTAAATATATTAAATTTATT 9021

†RNA

Qy 913 aagtcgtttactaatagtcattttgtcgcatctacttaataatagat---a 969

Db 17744 TATCTATTAAATTTATAAATAGTATATAGTTTTTTTTTTAAAAAAAATATTTTTTTTAA 17685

Qy 970 aattaattggtgacattagatcaagaacaaactagattttgtcccatctattgttaa 1029

Db 17684 AAAATTTTTTTTTTAAAAATGAAAAATAAATAATATATTTCATTATAAAATTTATTAT 17625

Qy 1030 aagctggtccgtttacattaaaataaggtacatgttcatgccagctataactatcgtgt 1089

Db 17624 TAAAAATTTTTTGTATTATTTTTTAAAAACATGATTTTATTATATAAATTTTTTATAA 17565

Qy 1090 tattctatcaatcacgctaatttttaacagtagaagaatgaatgtaatttttaaatagaag 1149

Db 17564 AAATAATCATTTTAAGAAATTTTTAAAAAATTTATATAAATTATTAAATAATTTAAT 17505

Qy 1150 ggtcaaatgttatttgatctaacacgtagggattaatttacttattt----tcctaaa 1204

Db 17504 TTTCTATATATATATATATATTATAAATTTCAATAATATATAAATTATAAATATATA 17445

Qy 1205 gaaataagtaaaatataattgaaatctaatacaaaaactttcatgatacttttatcata 1264

Db 17444 TAATTAATTAATTTATTAATAAAAAAAAAAAAAAAAAAATAATTTTTTATTATTAATTA 17385

Qy 1265 ttttacttataatttaattgtgagagtaacaaartaaaaaacatagaaacacccaaaa 1324

Db 17384 TATTAGTAATAAAATAAATTTTATTAAATTATTAGTTATTAAATTTATTATTATTAATA 17325

Qy 1325 gttagttatggtgtgactcatatacacagtgtaaaattgataaaattttttcttcgtca 1384

Db 17324 AATATTANTAATG-----AATAGATTTTATTTAATTATTAAATTATTAAATTTATTATTA 17270

Qy 1385 ttaattccatcatgggttttttttttctagtgtaagccataattatcaaaataatcatca 1444

Db 17269 TAAAAATTTAAAAATTTTTCATTTTAAATATATATATATATATATATAAATTTTAA 17210

Qy 1445 ttaatcctatcaataccccgcctgctccctccctcaataacttaaaccca--actaaca 1502

Db 17209 TTAATTATTTTAAATAAATTTTATTATAAAATAAATTTATTATAAAAAATAGTTTATTAAGT 17150

Qy 1503 cccagcaccaaacgcactttaatagccacctatttctagccatgctccttgcaactaaaga 1562

Db 17149 ATAATTTAATAAATAAATTTTTTTTTTAAAAAAAATATTTTTTAAAGTTTAAATATATAA 17090

Qy 1563 aaagtaagctaacctgcaatcattccatcatcgaggcctcacagataaagttggttgat 1622

Db 17089 TAAATTTATGAATAGGGGAATAAATTTATTTCATTTTACATATATATATATATATA 17030

Qy 1623 gggtttgaccaagttgttaaaacccgcctccctcaacttcccttttctttcatcctcccc 1682

Db 17029 TATATACAATTAATTAATTCAGATTTAGTGATTAATAAATAAATTTTATTATTAATCT---- 16972

Qy 1683 actccacacctccaattttcttcatatggttctattataagttctttataatcacagaa 1742

Db 16973 --TATATAATTTAATGAAAAATAAATATATGTATATATATAAATATATGAATTGAA 16916

Qy 1743 tcaagataagtcctcagcaacaaaaaacatggctctcgcgcaagatctgagctagtea 1802

Db 16915 TTTTATAAAAAATCATTTTAAATTTTATTATATTAATAAATTTTATTAAATTTATTA 16856

Qy 1803 gagctctgaatttggatcattattacagtgcaaaacagtgtaacaaaagctgttgagat 1862

Db 16855 AAATAATTTTATTATAAATAAATTTATTATAAATAAGTTTATTAAGTATAAATTAATA 16796

Qy 1863 aaacactgaaatctgctatag----ttgtttttggtttacatatgttccacgtgaaact 1917

Db 16795 AATCATTTTTTTTTTAAAAAAAATAATTTTTTAAAGTTTAAATATACAATAAATTTATGA 16736

Qy 1918 atgaagcatctctaagaaaacccaaactatcatatcaaccatcgatcaatgaatcgatt 1977

Db 16735 ATAGGGGAATAAATTTATTTTCATTTTTTTATATATATATATATATAAATTAATA 16676

Qy 1978 tcaattttcgcagtataaagttcccttttaacccctttcttttacttcattttataacgaat 2037

[illegible]

DMU37541/c
LOCUS DMU37541 19517 bp DNA circular INV 04-APR-2000
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION U37541
VERSION U37541.1 GI:1166529
KEYWORDS .
SOURCE Drosophila melanogaster.
ORGANISM Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 12511 to 12682)
AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and
Wolstenholme,D.R.
TITLE Drosophila mitochondrial DNA: a novel gene order
JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE 83090428
REFERENCE 2 (bases 5269 to 5695)
AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
flanking sequences and comparisons to mammalian mitochondrial tRNA
genes
JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE 83220794
REFERENCE 3 (bases 404 to 5272)
AUTHORS de Bruijn,M.H.
TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and
genetic code
JOURNAL Nature 304 (5923), 234-241 (1983)
MEDLINE 83245048
REFERENCE 4 (bases 804 to 1778)
AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.
TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in
Drosophila melanogaster and its sibling species
JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE 88174373
REFERENCE 5 (bases 5268 to 13619)
AUTHORS Garesse,R.
TITLE Drosophila melanogaster mitochondrial DNA: gene organization and
evolutionary considerations
JOURNAL Genetics 118 (4), 649-663 (1988)
MEDLINE 88212147
REFERENCE 6 (bases 441 to 2967)
AUTHORS Satta,Y. and Takahata,N.

c	25	134.8	4.4	130117	39	AC004907	AC004907 Homo sapi
c	26	133.8	4.4	158116	60	AC006281	AC006281 Plasmodiu
c	27	133.2	4.4	2426	8	SDU49822	U49822 Saccharomyc
	28	133	4.4	178087	39	AC005089	AC005089 Homo sapi
c	29	132.2	4.3	106650	39	AC007708	AC007708 Homo sapi
c	30	132.2	4.3	176552	39	AC004617	AC004617 Homo sapi
	31	132.2	4.3	203519	40	CNS01RHQ	AL162191 Homo sapi
c	32	131.8	4.3	321003	31	PFMAL4P3	AL035476 Plasmodiu
	33	131.2	4.3	80518	31	PFMAL13PA	AL109815 Plasmodiu
	34	130.6	4.3	170125	41	AC007465	AC007465 Homo sapi
	35	130	4.3	321003	31	PFMAL4P3	AL035476 Plasmodiu
	36	129.6	4.3	282806	60	AC006279	AC006279 Plasmodiu
	37	129.4	4.2	158116	60	AC006281	AC006281 Plasmodiu
c	38	129.2	4.2	224448	31	PFMAL4P4	AL035477 Plasmodiu
c	39	129	4.2	318221	31	PFMAL13P3	AL049184 Plasmodiu
	40	128.8	4.2	106650	39	AC007708	AC007708 Homo sapi
	41	128.8	4.2	176552	39	AC004617	AC004617 Homo sapi
c	42	128.6	4.2	12029	34	AE001372	AE001372 Plasmodiu
	43	127.6	4.2	75076	39	AC004948	AC004948 Homo sapi
c	44	127.6	4.2	204652	31	PFMAL13P6	AL049183 Plasmodiu
c	45	127.2	4.2	12029	34	AE001373	AE001373 Plasmodiu

ALIGNMENTS

RESULT 1

S79308

LOCUS S79308 913 bp mRNA PLN 30-NOV-1995

DEFINITION Rac13=21.8 kda GTP-binding protein [Gossypium hirsutum=cotton plants, cv. Acala SJ-2, boll fibers, mRNA Partial, 913 nt].

ACCESSION S79308

VERSION S79308.1 GI:1087110

KEYWORDS

SOURCE upland cotton boll fibers cv. Acala SJ-2.

ORGANISM Gossypium hirsutum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

REFERENCE 1 (bases 1 to 913)

AUTHORS Delmer,D.P., Pear,J.R., Andrawis,A. and Stalker,D.M.

TITLE Genes encoding small GTP-binding proteins analogous to mammalian rac are preferentially expressed in developing cotton fibers

JOURNAL Mol. Gen. Genet. 248 (1), 43-51 (1995)

MEDLINE 95379748

REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 170155] from the original journal article. This sequence comes from Fig. 1A.

FEATURES Location/Qualifiers

source

1. .913

/organism="Gossypium hirsutum"

/db_xref="taxon:3635"

gene

12. .602

/gene="Rac13"

/note="21.8 kda GTP-binding protein"

CDS

12. .602

/gene="Rac13"

/note="21.8 kda GTP-binding protein; pea Rho1 protein homolog/mammalian rac protein homolog; ; This sequence comes from Fig. 1A"

/codon_start=1

/protein_id="AAB35093.1"

/db_xref="GI:1087111"

/translation="MSTARFIKCVTVGDGAVGKTCMLISYTSNTFTDVPYVTFDNFS

ANVVVDGSTVNLGLWDTAGQEDYNRLRPLSYRGADVLLAFSLISKASYENIYKKWIP

ELRHYAHNPVVLVGTQLDLRDDKQFLIDHFGATPISTSQGEELKMGIAVTVIECSS

KTQQNVKAVFDAAIKVALRPPKPKRKPCKRRRTCAFL"

BASE COUNT 307 a 169 c 172 g 265 t

ORIGIN

Query Match 10.0%; Score 304.8; DB 8; Length 913;

Best Local Similarity 97.8%; Pred. No. 8.5e-24;

Matches 309; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1806 cctcgaatattggatcattattacagtcacaaacagttacaaaagctgttcagataaa 1865

Db 597 CTTTGAATATTGGATCATTATTACAGTCAAAACAGTTAACAAAAGCTGTTGCAGATAAA 656

Qy 1866 cactgaatctgctatagttgtttgtttgtttacatattgtccacgtgaaactatgaagca 1925

Db 657 CACTGAATCTGCTATAGTTTGTGTTTGTGTTTACATATGTTCCACGTGAAACTATGAAGCA 716

Qy 1926 tctctaagaaaacccaaactatcatatcaaccatcgatcaatgaatcgatttcaatttt 1985

Db 717 TCTCTAAGAAAACCCAAACTATCATATCAACCATCGATCAATGAATCGATTTCATTTT 776

Qy 1986 cgcagtataagttccttttaactcctttcttttacttcattttataacgaattctatgga 2045

Db 777 CGCAGTATAAGTTCCTTTTAACTCTTCTTTTACTTCTTTTATAACGAATTCATGGA 836

Qy 2046 taatgttcctacaaacatgtcattacaatgtttaattataaattccattcttctatgtt 2105

Db 837 TAATGTTCCCTACAACATGTCATTACAATGTTTAATTATAAATTCATTCTTCTATTTT 896

Qy 2106 actaagatattagtaa 2121

Db 897 ACTAAAAAATAAATAA 912

RESULT 2

DMU11584/c

LOCUS DMU11584 4601 bp DNA INV 23-JUL-1994

DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.

ACCESSION U11584

VERSION U11584.1 GI:508826

KEYWORDS mitochondrial DNA; A+T region; tandem repeats.

SOURCE fruit fly.

ORGANISM Mitochondrion Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 4601)

AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.

TITLE Sequence, Organization and Evolution of the A+T Region of Drosophila melanogaster Mitochondrial DNA

JOURNAL Mol. Biol. Evol. 11, 523-538 (1994)

MEDLINE 94285822

REFERENCE 2 (bases 1 to 4601)

AUTHORS Kaguni,L.S.

TITLE Direct Submission

JOURNAL Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of Biochemistry, Michigan State University, East Lansing, MI, 48824-1318, USA

FEATURES Location/Qualifiers

source

1. .4601

/organism="Drosophila melanogaster"

/organelle="mitochondrion"

/strain="Oregon-R"

/db_xref="taxon:7227"

/dev_stage="embryo"

gene

1. .4601

/gene="mt:ori"

/note="mitochondrial origin"

/allele=""

/db_xref="FlyBase:FBgn0013687"

repeat_unit

650. .1022

/gene="mt:ori"

/note="repeat I-A"

/db_xref="FlyBase:FBgn0013687"

/rpt_type=tandem

repeat_unit

1023. .1360

/gene="mt:ori"

/note="repeat I-B1"

/db_xref="FlyBase:FBgn0013687"

/rpt_type=tandem

Search completed: September 2, 2000, 22:57:50
Job time: 19286 sec

Qy 5285 aattatcttaactaaaatctaaaattttattttaacctatt 5324
: :||| | | | | | | :|:| | : |:|
Db 1161 WTWATTATTATTTTATATTATATATWWTATWWTATWWT 1200

Oy 3609 caacatcgtatatttacttattaatacataatttatcataattttatggaattgagacc 3668

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	Location/Qualifiers				
source	1..1101				
		/organism="Drosophila melanogaster"			
		/db_xref="taxon:7227"			
		/clone_lib="RPCI-98"			
		/clone="BACR29P01"			
		/note="end : TET3"			
BASE COUNT	366 a	66 c	104 g	351 t	214 others
ORIGIN					

Query Match 2.1%; Score 115.6; DB 122; Length 1101;
Best Local Similarity 40.6%; Pred. No. 2.3e-08;
Matches 231; Conservative 81; Mismatches 255; Indels 2; Gaps 1;

Qy 4703 atagaaattctaaatggttatagtgttatagtgttagtgaaattaatttt 4762
|:| ::||| |:| ||||| :| :| | :|||
Db 969 AYATTTWTATACATAATWTTTTTTTATATACAATTTAAAAATAAAAWTAACWTAATT 910

Qy 4763 aaatgttgtagtctaagtgttaacatcacttgcgctgatttatgttatgttatgtattttac 4822
| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 909 AWAANAACATTNTTTTAAWNTAATTTTAACTACAWTNTAAWNTAAAAAATTTTAAAWTAA 850

Qy 4823 tttaatgatattgcatgtattgttaatttaacattgcttgatcattatactcttctacta 4882
: || ::||| | | : :| | :| :| |||:: : | | ||
Db 849 AAAAAAAAAWATAAATTATATWATATWATAAAWTTATAWATTATTMW--WAAWTTATA 792

Qy 4883 ttaattataaatggcactgttttgtttaaactttttacaagttaagacatgtataaatat 4942
: | | | : : : | | : : : | | | : | | | : | | |
Db 791 WNAATTTTTTTTATAAANWTTTATWATTAATTTTAAWTTTAAATAAAAAAATTTAT 732

Qy 4943 atgacaatataattacaggttttagttcaatgttagctactttagtatgttattgatgat 5002
| : | ||: | ||||: | |: | :: | | : | | :|| | | :
Db 731 TTTTATTWTWTATWAAAWTTTTWTTTTWAATTWYTTTAATAWTAAAATWTTAAAWAW 672

Qy 5003 ctttaattacattttaaacaaatttccacttaaaatttttaataaataaacaataattatt 5062
| :||| | : | | :||| | || |:| :| :| ||
Db 671 ttaawttaaataaattattttaaaatttwaawttataaatttawadttataawttttttt 612

Qy 5063 gtaatataacattaaatgcaacaaaaaatgaataaataaaaataagcaataat 5122
 : : | : ||| | | ||| : | ||| : | | | : || : | : : |
Db 611 HNCWTTWAATATATAAAATWAAAAATTAAATTWTTCCTTATATWWTAATAAAMATWWWTT 552

Qy 5123 tgttataaatattgtaataataatgtaccatattcttaactgaataagggtctaacctat 5182
 |||||:: || |:| ||| || | | |
 Db 551 ATTATAAWWAATTATTAAWWTGATTTTATTATTATTTTATTTTAAAWTTTATTATA 492

Qy 5183 aatccctaaaatttcagttttaaattttttatacctaccatatatttagaacctcttttta 5242
:
| | : | | | | | | | | | | : | : | | | | | : | : | | :
Db 491 ttaatttatataattttatattatgtttttttttatataccttcmaaatattttmccccoytttw 432

Qy 5243 aatatattaaaaattttaattataccaatt 5271
|| || |||| : || | ||
Db 431 TTATAGTTTTTTTTTTCCTCTNCNTT 403

```

RESULT 10
B11102/c
LOCUS      B11102      1187 bp      DNA      GSS      14-MAY-1997.
DEFINITION F19C22-r7 IGF Arabidopsis thaliana genomic clone F19C22,
            genomic survey sequence.
ACCESSION  B11102

```

```

VERSION      B11102.1  GI:2092386
KEYWORDS     GSS.
SOURCE       thale cress.
ORGANISM     Arabidopsis thaliana
              Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
              Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
              Brassicaceae; Arabidopsis.
REFERENCE    1 (bases 1 to 1187)
AUTHORS      Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
              Ecker,J.
TITLE        BAC End Sequences at ATGC
JOURNAL      Unpublished (1997)
COMMENT      On Sep 10, 1998 this sequence version replaced gi:3556525.
              Other_GSSs: F19C22-Sp6
              Contact: Ecker J.
              Arabidopsis Thaliana Genome Center
              University of Pennsylvania
              Dept. of Biology, University of Pennsylvania, Philadelphia, PA
              19104
              Tel: 215-898-9384
              Fax: 215-898-8780
              Email: jecker@atgenome.bio.upenn.edu
              Seq primer: T7
              Class: BAC ends
              High quality sequence start: 72
              High quality sequence stoo: 353.

```

FEATURES	Location/Qualifiers
source	1..1187
	/organism="Arabidopsis thaliana"
	/strain="Columbia"
	/db_xref="taxon:3702"
	/clone="F19C22"
	/clone_lib="IGF"
	/sex="hermaphrodite"
	/note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
	Produced by Thomas Altmann"
BASE COUNT	385 a 51 c 60 g 594 t 97 others
ORIGIN	

Query Match 2.1%; Score 115.2; DB 120; Length 1187;
Best Local Similarity 47.0%; Pred. No. 2.7e-08;
Matches 362: Conservative 0; Mismatches 402; Indels 6; Gaps 2;

Oy 3208 ttatattacggaatgtaatatattatattttaaaataaaattatgttattagattcttaat 3267
| | | | | | | | | | | | | | | | | | | | | |
Dh 1185 TAATATTATTTTTTATAAATTATAAAAAATTAATATATATATATAAAAAATAAAATTT 1126

Qy 3268 attttggagcattccatactataatttcgtaacataatatataaaatagtaataataaag 3327
|| || || || || || || || || || || || || || || ||
Db 1125 TATTANAATAATATAATAATAAATAATTATTATATATAATATAAAATTTAAAAATAAAA 1066

Qy 3328 tgtaattaacttttaaattacaagcataatattaaatttgaatcaattaaatttttatctc 3387
| | | | | | | | | | | | | | | | | |
Db 1065 TATAATTAAAANTAAAAATATAAATTAATATAAATATAAAATATAAATAAAAATAAAAAT 1006

Qy 3388 tattattttaattaatttagtctatttttcaaaaataaaatttaaactcaataaaaaata 3447
||| ||| | | | | | | | | | | | | | | | |

Dh 1005 TATAAATTAAAAATATTAATATTATAAAATATTTTATTAAATANATATATTAAAAAAAAAAT 946

Qy 3448 atttttccctaatgttgaaacaactcatgttatacttcaaaattataagttattatatta 3507
 ||| ||| | | | | | | | | | |
 Db 945 TATTTTATATAAAAAAANAANNAATATTTTTATAAATATTAAAAATAAAATAAAA 886

Qy 3508 ccttgatgattattatttagtatattaattctgattataattatgggtggagatacaatcg 3567
 | | | | | | | | | | | | | | | | | | | | | |
 Db 885 AAAATTTTTTAAATAAATAATATAAAAAAATAAANAANAATAAANAATAAATAATAT 826

Qy 3568 cttccactaaaatttttaactatgatttataaaatttttcaacatcgta--tatttac 3625
 || | | | | | | | | | | | | | | | | | | | |
 Db 825 TTTAAAAAATTAAATATTTAAAAATATTTAAAAAATATTTAAAAAATATATTATNANA 766


```

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
    source            1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR08K10"
                      /note="end : TET3"
BASE COUNT          201 a    64 c    131 g    202 t    503 others
ORIGIN
```

Query Match 2.3%; Score 126.2; DB 122; Length 1101;
Best Local Similarity 21.5%; Pred. No. 6.2e-10;
Matches 151; Conservative 310; Mismatches 238; Indels 4; Gaps 2;

Qy 3299 acataataatataagataagtaataaactttaattacaagcataatat 3358
| | : : : : | | : | : | | : : : | | : | : | | :
Db 398 ATAWWWWWTTTTTTTAWAAWAAAAATAATTWAAWAAAAAATTTWAAWAAAAAW 457

Qy 3359 taaatttgaatcaattaattttttatttctattattttaatttagtctatttttc 3418
: |:| |:| || || |||: || |:| |:| : : : : |||:
Db 458 AWTAWNTTAWTAAAAAAAAAAATTTTTTTTTTTWTWTATATWTTTWTWTAAAW 517

Qy 3419 aaaataaaatttaaatctaaataaaaaataatttttccctaagtgtgaacaaactcatgtt 3478
||| ||| :||| |:|||| |:|:|:|:|: |:|:| |:|:| :|:| ||
Db 518 AAAAAAAAAAAWAAWAAWATAAATTTTWWTTTYYTAAWATAAAMCAAWYYHYTTT 577

Qy 3479 at-acttcaaaattataagttattatattaccttgatgattttattagtattattaat 3537
| ::::: |::: |:::|::: : :::::| : ::::|
Db 578 YTYHYTYTYTWTWHYHTMTTAAWAHTTWHYHTYAMWHTWTWHWTWAAWHTTYYTA 637

Qy 3538 tctgattataattatgggtgggaacacaactcgctttccactaaatatTTTAACTATGATTTA 3597
:: :| :: : | |:| :: :||: ||::| :|||| : : :| : : :|
Db 638 AYYYYYTCYYYHYMHWHHAHAAAATTTTTTWTTHAYHHWATYYHHHHMYCAMCMCTHT 697

Qy 3598 taaatttattttcaacatcgtatatattacttattaatacataatttatcataattttatgg 3657
:: :: :: |: : ::|::: :: :: :: || |: :::: : :
Db 698 CHHCYYYHHYTAHHTTHHHNYAHYMYWYYWYMYMCTACTYHHHHHHHHYHAHTTW 757

Qy 3658 aaattgagaccaagaacaacattaagagaacaaattctatacaagaacaaatttagaaaaaa 3717
| : : : : | | | : | | : : : || : : : : : : : : : :
Dh 758 YAAAHAMWGNHAYAAAAAATTTHTYHTHTHYMHTMYTHYMYTCCSYMTYHCN 817

Qy 3718 atgtactcttttagtgtaatttttaagtactcttaaccaaacacaaaaattccaatcaaatgaa 3777
: ||: |: ::::|::: : : : : : : : : | : || :
Dh 818 YYHAYTCCTWTHTHWWWTATHTWYHHWHTTTTWWAWHTHTGCGGWTATGWWATGCG 877

[illegible]

Qy 3838 tattatgaaaaataatcttattactcgaactaaatgttgtcacaaattatta---tot 3894
::: : : | : : ::::: : : : | : : | : : ||:
Db 938 nnnmmatmtttmtnnmmssnnbnncnmynnmymuuscnhsucstnathtnyumtsy 997

Qy 3895 aaataaagaaaaaacacttaattttttataacattttttcatatattttgaagattatatt 3954
: |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: : |: :
Db 998 HVCWHTYTYTAYWAAWTAAHMTTATATTTTTTTTTTAAHATATTTTTTTTTTAAWACTCHHW 1057

Qy 3955 tgtatatattacgtaaaaatatttgacatagattgagcacottc 3997
: |: :::: ::::|:: : :: ::: | : ::|:
Db 1058 YHTHCTWYYHHTYHMMWAWMMWAWHWHMMYAHYHWAHHCWYYTM 1100

```

RESULT      4
CNS0167M/c
LOCUS       CNS0167M      1201 bp      DNA              GSS              26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
             BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION   AL106396
VERSION     AL106396.1  GI:5621701

```

KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM *Drosophila melanogaster*
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; *Drosophila*.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European *Drosophila* Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> -. This *Drosophila melanogaster* BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

```

FEATURES             Location/Qualifiers
    source            1. .1201
                        /organism="Drosophila melanogaster"
                        /plasmid="pBelobAC11"
                        /db_xref="taxon:7227"
                        /clone_lib="DrosBAC"
                        /clone="BACN15M24"
                        /note="end : T7"
BASE COUNT           323 a      87 c      79 g      551 t      161 others
ORIGIN

```

Query Match 2.2%; Score 121.6; DB 123; Length 1201;
Best Local Similarity 37.4%; Pred. No. 3e-09;
Matches 280; Conservative 101; Mismatches 368; Indels 0; Gaps 0;

Qy 3219 aatgtaatatattatatttttaataaaattatgttatttagattottaatatattttggagca 3278
:|: |: ||| | || |: : || |: | | |
Db 1193 WATAWAWATATATAAATATAAAAAATAAWAWAATATATAWNAAAAANTATAAAAAWA 1134

Qy 3279 ttccatactataatttcgtaacataattataaaatagtaataaagtgaattaact 3338
 ||: ||||| : || ::||| ||| : | | :|| || ||
 Db 1133 AAAATATATATATAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAA 1074

Oy 3339 ttaaataccagcataatattaaatttgaatcaattaatttttatctatttttaa 3398
:: || | :|| : | || | |||| || :| :|
Dh 1073 aawwtthtnanaattttttttttatawaattttttttttttttatawawaaaaa 1014

Qy 3399 ttaatttagtctattttttcaaaataaaatttaaatctaaataaaaaataatttttccta 3458
|| || || | : | : || || || | : : : || :
Db 1013 AAAAAAAAAATTTAAAAATAAAWTAAATATWAAAAATTTTAAAAATTTTWTWTTTTW 954

Qy 3459 atgttgaaacaactcatgttatacttcaaaattataagtattattacccttgatgatt 3518
| ||| :| : :| || :||| :| :| ||| :|||
Db 953 AAAAAAAAAAATAATAAAATTTTTTTTATATATAAAWAATTTTTTWTAAAAAA 894

Qy 3519 tatttattagtagtataattaattctgattataaattatggtgggatacaatcgctttccactaa 3578
| : | : | | | | | | | | | : : | :
Dh 893 AAAATTA AAAATTTAAAAATATATAAATAAAAAWAAAAAAAAAAAAAAAAAAAAAAWAAAAATTT 834

Qy 3579 atattttaactatgattataaattttatttcaacatcgatatatttacttattaatacata 3638
 :||| ||| | :| ||| | | :|| | | | :| :|
 Dh 833 TWAATTATAAAATTTTAAAAAATTTTAAATTTTNNAAATAAAAAATAAAAA 774

Qy 3639 attatcataattttatggaaattgagaccagaagaacattaagagacaacaaattctataac 3698
::: | | ::| || ||::|| | | |||| | : | ::|| | |
ph 773 WWWAAAAAATWTTAAATATAAATTAAATNCATAAAACAAAAAAATTTGHTTAAAAAATA 714

Oy 3699 aaagacaatttagaaaaaatgtacttttaggtaattttaagtactcttaaccaaacaca 3758
 ||| | | | ||||| | ||| : | : :: : | : |: |:

Nh 713 AAAAAAAAAAAAAAANAATTTTTTACAAAMMAMMMVMMMMCAAAAAMMAAVVV 654

Ov 3759 aaaattcgaatcaaatgaactaaataagataatataacatacgaacatcttacttqtaa 3818

Db 813 ATAWATWAAAWAWATAWATWATATAWWAATAWAWAAAAATWTAATATWAATWATAA 872

```

RESULT 1
CNS00EVL
LOCUS      CNS00EVL      1101 bp      DNA      GSS      04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Qy 3220 atgtaatatattatatttttaaataaaaattgttatttagattcttaaatattttggagcat 3279
| | | | | | | | | | | | | | | | | | | | |
Db 3932 AAGGATGATAATACTTCAAAAAATCATGAGCCAACCTTAGATATTCTCTTTTTAACATT 3873

Qy 3280 tccactactataatttcgtaacataattttaaaatagtaataataagtgtaattaacct 3339
| | | | | | | | | | | | | | | | | | | | |
Db 3872 CCATCATTTTTTTTATCACACTTTTTAGTTCATAAAACTTAAGACCATTATTTTATGT 3813

Qy 3340 taaattacaagcataaattataaattttgaatcaattatttttatctattatttttaat 3399
| | | | | | | | | | | | | | | | | | | | |
Db 3812 TTTGTAATTTTAAAATAATCACATTTTGGTTTATTCTTTTTATCGATAATATT---GG 3756

Qy 3400 taatttagtctattttttcaaataaaaaattaaactaataaaaaataattttccttaa 3459
| | | | | | | | | | | | | | | | | | | | |
Db 3755 TGGATTGTCTATTTTTTTAGGAATTCATTTTTATAGTATTATCACTTTTTGTTTTAT 3696

Qy 3460 tgttgaaccaactcatgttatacttcaaattataagtattatattacccttgatgatt 3519
| | | | | | | | | | | | | | | | | | | | |
Db 3695 TCATAATTATTTTGAAAAATAGTAAATACCGTGTAAATATACAACCTAAAAATGTTATTA 3636

Qy 3520 atttattagatattataattctgattataattatggtgggatacaatcgctttccactaaa 3579
| | | | | | | | | | | | | | | | | | | | |
Db 3635 ACTTTTAAGTTTTTTTTTTTTTTTTTTTTTTTATATTAAGAATAAATGTAAACCTTAAA 3576

Qy 3580 tattttaactatgatttataaattattccaacatcgatatattacttattaatacataa 3639
| | | | | | | | | | | | | | | | | | | | |
Db 3575 TATTGGAGTATAAATAATATATATATTATAAC--GAGACAATTAGTTAAAAAAAATAG 3518

Qy 3640 ttatcataattttatggaaattgagaccaagaacattaagagaacaaattctataaca 3699
| | | | | | | | | | | | | | | | | | | | |
Db 3517 TTAAAAAAAATCGTAAAAAAAATATGAAAAATAATGGATATATAATTGAATGAATAA 3458

Qy 3700 aagacaatttagaaaaaaatgtacttttaggtaattttaagtactcttaacccaacacaa 3759
| | | | | | | | | | | | | | | | | | | | |
Db 3457 CATAAAA---AGATGACAATTTATCAAACGTTAATTAAAAATAACTTAATCATACAAAA 3401

Qy 3760 aaattcaaataaaatgaactaaataagataataataacatacggaacatcttacttgtaat 3819
| | | | | | | | | | | | | | | | | | | | |
Db 3400 AAAAAGGACAAAAACAGGAAAAAGGAATAAAGTGTAAGAGAACACAAACATTTAAAGA 3341

Qy 3120 actgctcacagataatgtttaaaatgaaattaaaataaggtagcctggctgcacacacaca 3179
| | | | | | | | | | | | | | | | | | | | | |
Db 1381 AGTACGAACATAAATATGTATAAACCAAAAAAATTGATTAAAGATAAAGTTATATGTTTG 1440

Qy 3180 aaaaaaacactaatgttggttggttgaatttttatattaccggaatgtaattatatatttttaa 3239
| | | | | | | | | | | | | | | | | | | | | |
Db 1441 ATTTAATAAAAATAGTTTAGTTTTAAAAATTTTATATCATTTTTTAAAAAATGAAAATGTTG 1500

Qy 3240 ataaaattatgttatttagattcttaatatatttggagcatccatactataatttcgtaa 3299
| | | | | | | | | | | | | | | | | | | | | |
Db 1501 AAAAAAAAAAATTTTTTTTTTTTTTTTCAACGGGACGATGTAATATCATATGATTCAAAT 1560

Qy 3300 cataatattaaaatatagtaataataaagtgtattaacttttaattacaagcataatatt 3359
| | | | | | | | | | | | | | | | | | | | | |
Db 1561 TAAAGTTATTAACAATAATGTAAAAATTATAAAAACTAACCTAGTTATAATTACTTTC 1620

Qy 3360 aaattttgaatcaattaattttt---attctattatttttaattattagttctattttt 3416
| | | | | | | | | | | | | | | | | | | | | |
Db 1621 CCTCTTTTTTTTTTTTTTTTTTGTGCATGACACTTTTTTTTTTGTGCATGACACTTTT 1680

Qy 3417 tcaaaataaaatttaaatctaataaaaaaataatttttccttaatgttgaacaactcatg 3476

Db 1681 TTTTAAAAAAAAAAAAAAAAAATGTTAAATACTATTGTGATGACATTCATTTTTCCTAGTT 1740

Qy 3477 ttatacttcaaaattataagttattatttaccttgatgattatttattagt--atatt 3534

Db 1741 TTTTTTTAGATAGATATAAAAAATAAATTGCCTATCGATATATACTTAATTATTATAAGATT 1800

Qy 3535 aattctgattataattatgggtgggatacaatcgctttccactaaatttttaactatgat 3594

Db 1801 GAATAATATTTTAAATTTTAAATAAATCTACTTTTTTTTTTTTTTCTTTTTTTTTTAAA 1860

Qy 3595 ttataaaatttatttccaacatcgatatattacttattaatacataatttatcataatttta 3654

Db 1861 TTTTAAATTTTTTTTTTTTATTAGATCTCATAATTAATAATCAATTAAAATTAAGT 1920

Qy 3655 tggaaattgagaccagaagaacattaagagaacaattctataacaagacaatttagaaa 3714

Db 1921 TATTTTTTAAATATGCAAAACTATAAAAACTAATGTAGTTTAACCAACTTTTTCTATT 1980

Qy 3715 aaaatgtacttttaggtaatttttaagtactottaacaaacacaaaaattcfaatcaaat 3774

Db 1981 TCTTTTTTTTTTTTTTTTTTTTTTACTTTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2040

Qy 3775 gaactaaataagataataataacatacggaacatcttacttgtaattctt-----acatt 3827

Db 2041 AAACCTCATTTATAAATTAATTAATTACTTGGTTTTTTTTGATTTTTTTTTTAATAAATT 2100

Qy 3828 cccataattttattatgaaaaaatacttataattactogaactaaattgtgtcacaaatt 3887

Db 2101 AAAATTTTATTCTCTATCTAATTATACCTTATTATAAATATTGGATAATATATCAAATA 2160

Qy 3888 attatctaaataaagaaaaacacttaatttttataacattttttcatattttgaaagat 3947

Db 2161 TTTTTCAGTTTGGCATGACAATTTTAATTATATTATTTTTTTTGATTAAATTTTTTTTT 2220

Qy 3948 tatattttgtatatatttacgtaaaaaattttgacat 3982

Db 2221 TTTTTTTTTTAAATTTCTTTTTTTTTTTTTTTTAT 2255

```

RESULT 10
US-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 1.8%; Score 100.4; DB 4; Length 19124;
Best Local Similarity 44.1%; Pred. No. 1.4e-09;
Matches 877; Conservative 0; Mismatches 1091; Indels 22; Gaps 10;

Qy 1970 attgaaacgtttaagaattttactactgcaaatcagaataagtgatttttttag 2029
Db 4660 ATTCAATATTAGAGATTATGTAATATTGTTATGTCGTAATATATATTAATTAATT 4719
Qy 2030 aaagattaaataagtttagtattacgatttttagtttgattgggaaagtaagtatgt 2089
Db 4720 GTTTTTTATGATGATGTTATCTAATAATATATTCATATGTCATAGTCATGCAATGA 4779
Qy 2090 ttttgacataaattatttgacaataattaagttttctagggaataaacggaaatatcttc 2149
Db 4780 ATATAAATATGTTATATTATATATTAATAAGTAACACAGA-ACATTA 4838
Qy 2150 tctctttttgttaaaattactaatgcaagaacaaacacggtttgggggcaataatctta 2209
Db 4839 TATATAGTAATAAAGAGAAATAATATATTTTATGTTATATATATTAGTTATTATA 4898
Qy 2210 gctttaagtagtcagtgtaactctcaaaactctgggcataactcttaggtgagttgtgtg 2269
Db 4899 AAGGGGAAATTCATAATATTGAAATTTTGTATATGATATAGTTAAGTTAAAA 4958
Qy 2270 tgctacagtagtaagtcctatagaaactcctgacaaaacacatgacgtcagggcgaa 2329
Db 4959 AAAAAAAAAAACAAGAAACAAATGGAAGCATAAAATGTTACTGTAATAGGATAAAA 5018
Qy 2330 tctacaacttttctctttctcaattacatattgattcaagttccgatctataat 2389
Db 5019 TATATTATATAAATGTTTATTTTATCTTAAAAAGGTTCTTATTATAACATTAATAAAAA 5078
Qy 2390 aatttattacgattttcaatttcaattaccttatcatctattataaataaagtc 2449
Db 5079 TTTGTCCTATTTTATAAATAAATTAACATACATTTACATAATGAAATTCGATTTTGTGTTT 5138
Qy 2450 gttcaactcagttttcgaaagttcccaaaattttgaattttataaattttccctaa 2509
Db 5139 TTTTGATGAATATTATGGACTAATTATTATATGTAATGCGTCTATATAAATAAATA 5198
Qy 2510 aaccgaaatagtttatctcttcaatttaagttttcatttttcaatccgattt-caatttc 2568
Db 5199 ATTTTATTTAAAAAATGAAAAAAGAAATAAATATCTGATTTTGTAGTTCAATAGC 5258
Qy 2569 atccttttataactctctattatctataattacataaatttcaattatatttgaatat 2628
Db 5259 TTAATATAATTATGGACTCATATATATATATATATCTTTACACAAGTAATAAGTAA 5318

Qy 2629 ttacacttttagtccttaagttcaaaactataaattttcacttttagaataatcattttt 2688
Db 5319 ATATTATTTTAACTTTAATAAGGAAAAATAAATAAATAAGAA---TACTGAATAA 5375
Qy 2689 cacatctaagcatcaaatttaaccaaatgacacaaatttcattgatttagatcaagct 2748
Db 5376 TAAGTCATATTATACATTTTTTAAAAATGTAACATAATTACAAATACGTAACTGATTA 5435
Qy 2749 tttgagtccttcaaacataaaaaattacaaaaaaacaaacttaaaatcatttatcaat 2808
Db 5436 TAGAAATAATAAGAATTAAATATTAAGGATAAATAAATAATTAAATTTATATTTTTT 5495
Qy 2809 ttgaacaacaagcttgccgcaatgtaagagcttaaaatgctctctttgtttctttt 2868
Db 5496 ATGTCAATTTATGTTATATTATATATTAACATGATTA-GTTTTTGAATAATATTA 5554
Qy 2869 tgttgcaaacggttgagagagaggggaaggaagattgaccatattttttattatgttt 2928
Db 5555 AATATCATATAATAATAAATAATAGTTAAATAATAGTATTTTATACAAATACATACT 5614
Qy 2929 taacataataattataaatttaataataataataataactttgttgatgagcagtgaggag 2988
Db 5615 TATAAGTATATCATATAATATATATATATATATATTTATGTTTGTGTTGGTGTAT 5674
Qy 2989 atacgtaaagtattttaacattataactttttgcaagcagttgctgtctaccgaagct 3048
Db 5675 ATAAGGCTATAGTATATATGTTGTTGTTTATATATTTATATGTAATAGATACATAT 5734
Qy 3049 gatcaaa-----gtttgagctgccttcaatgagccaaattttgcccataatggataaa 3101
Db 5735 AAGTTAATATATTTTGTGTATATGCTGTGTTAAGATAGATATGCATTACAGTTAAG 5794
Qy 3102 ggcaattttgttttagttcaactgctcacagaataatgtttaaataaataaataaaggtg 3161
Db 5795 GGTATAGTTTTTTTTTTTTTTTTTTGTACATATATATAAAAAATAGATAACTAACAATA 5854
Qy 3162 gcctgtgcac 3221
Db 5855 TGCATATTACAAGAATAATATTGTATAAATATATATATATATATATATAAAGACAT 5914
Qy 3222 gtaattattatatttttaaaataaaattatgttatttagattcttaatttttgagcattc 3281
Db 5915 -TAAACTATACATAAGTAATAGTTTATTATATATCATCCTTTTATTATTATAATTTT 5973
Qy 3282 catactataatttcgtaacataattataaaatagtaataaagtgtaattaacttta 3341
Db 5974 TTTTGTCTTCTGCTGCTCTCTTTTGTGTTTATAATATAACAATATAAACAATA 6033
Qy 3342 aattacaagcataatatttaattttgaatcaattattttttattctatttttaatta 3401
Db 6034 TCAGTATTGGAATATAAATAAATTTTCTACATATATGCATATATATATATATATA 6093
Qy 3402 atttagctattttttcaaaataaaatttaaatcaataaaataaattttt---cctta 3458
Db 6094 TAT 6153
Qy 3459 atgttgaaacaactcatgtttatacttcaaaattataagttattattttacottgatgatt 3518
Db 6154 ATTTTATATATTTTAGTATATATCTTTAAGATATTATTAATATTATATAGTAGCATAT 6213
Qy 3519 tatttattagttatataattctgattataattatggtgggatacaatcgctttccactaa 3578
Db 6214 ATGTATTATATATAAATAAATTTTCAATTATATAAATATATAAGAACATGAACATTTT 6273
Qy 3579 atattttaactgattt-tataaatttttcaacatcgatatatttacttattaatacat 3637
Db 6274 ATTAATAACTCATATTTGAATATATATTTTAAATGTGATTTTACTTATTTTAT 6333
Qy 3638 aatttatcataattttatggaattgagaccaagaacattagagaacaaattctataa 3697
Db 6334 ATTATACAATAAATTTTGAATTCATAAATGCATGAATACATAAATAAATAACAACAA 6393

Qy 4252 gcagcggctcgacgtttatttcgagacacaaagcaacctcatcagagctccacaaattggct 4311
|||||
Db 133 TCAGCGGCTCGACATTATTTCAGACACAAACCACTCATCAGAGCTGCCACAATTGGCT 192

Qy 4312 tcaaaatagcaaaagcagagagctgtgaatcagaaaagccagaatacaaacagccaaagta 4371
|||||
Db 193 TCAAAATACGAAA-----AGCACAAAGAGTCTGAATACAAACCAACCAATA 239

Qy 4372 tcacgaagagtactcaaaacttgagaagcctgaaatgcaaaaggaggaacaaacaaac 4431
|||||
Db 240 TCACGAAAGTACCCAAACATGAGAGCCTAAATGCACAGGAGGAAAAACAAAC 299

Qy 4432 ctgcaaacagcatgaagaggtaccacaggtcacacgaatcaaaaggagcaaaagagtagca 4491
|||||
Db 300 CTGCAACATCATGAGAGTACCACGAGTCACGGAATCGAAGGAGCAGCAAGAGTAGCA 359

Qy 4492 gaaagaaaatctcga 4506
|||||
Db 360 TAAAGAAAAACCGA 374

RESULT 5

US-08-787-335-18

; Sequence 18, Application US/08787335

; Patent No. 5981834

; GENERAL INFORMATION:

; APPLICANT: John, Maliyakal E.

; APPLICANT: Umbeck, Paul F.

; APPLICANT: Brill, Winston J.

; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS

; TITLE OF INVENTION: FOR ALTERED FIBER

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles and Brady

; STREET: P.O BOX 2113

; STREET: FIRST WISCONSIN PLAZA

; CITY: MADISON

; STATE: WISCONSIN

; COUNTRY: U.S.A.

; ZIP: 53701

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: Macintosh

; SOFTWARE: Microsoft Word 4.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/787,335

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/530,797

; FILING DATE:

; APPLICATION NUMBER: US 07/253,243

; FILING DATE: 04-OCT-88

; ATTORNEY/AGENT INFORMATION:

; NAME: Nicholas J. Seay

; REGISTRATION NUMBER: 27,386

; REFERENCE/DOCKET NUMBER: 1122990245

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1283 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; HYPOTHETICAL: no

; ANTI-SENSE: no

; ORIGINAL SOURCE:

; ORGANISM: Gossypium hirsutum

; STRAIN: Coker 312

; DEVELOPMENTAL STAGE: 15 day old fiber cells

; TISSUE TYPE: fiber cells

; IMMEDIATE SOURCE:

; LIBRARY: CRFB15

; CLONE: E9

US-08-787-335-18

Query Match 4.8%; Score 265.4; DB 4; Length 1283;
Best Local Similarity 84.3%; Pred. No. 9.4e-39;
Matches 316; Conservative 0; Mismatches 46; Indels 13; Gaps 1;

Qy 4132 aatacacgtttctttcttctatttgattacacacgtcattcgcaccccttc 4191
|||||
Db 13 ACTAAATTCCTTCTGCTTCTATTTTGTAACCATGGCTCATACCTTTTGTCTCTTC 72

Qy 4192 ttccttttccaaacttttactcataagtgtctcactagtacggtagccactgtttcg 4251
|||||
Db 73 TTCCTTTTCCAACCTTTTACTCATTACTGTCTCACTAATAATCGGTAGTCACACCGTCTCG 132

Qy 4252 gcagcggctcgacgtttatttcgagacacaaagcaacctcatcagagctccacaaattggct 4311
|||||
Db 133 TCAGCGGCTCGACATTATTTCAGACACAAACCACTCATCAGAGTGCCACAATTGGCT 192

Qy 4312 tcaaaatagcaaaagcagagagctgtgaatcagaaaagccagaatacaaacagccaaagta 4371
|||||
Db 193 TCAAAATACGAAA-----AGCACAAAGAGTCTGAATACAAACCAACCAATA 239

Qy 4372 tcacgaagagtactcaaaacttgagaagcctgaaatgcaaaaggaggaacaaacaaac 4431
|||||
Db 240 TCACGAAAGTACCCAAACATGAGAGCCTAAATGCACAGGAGGAAAAACAAAC 299

Qy 4432 ctgcaaacagcatgaagaggtaccacaggtcacacgaatcaaaaggagcaaaagagtagca 4491
|||||
Db 300 CTGCAACATCATGAGAGTACCACGAGTCACGGAATCGAAGGAGCAGCAAGAGTAGCA 359

Qy 4492 gaaagaaaatctcga 4506
|||||

Db 360 TAAAGAAAAACCGA 374

RESULT 6

US-08-487-826B-13/c

; Sequence 13, Application US/08487826B

; Patent No. 5993827

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; APPLICANT: Chitnis, Chetan

; APPLICANT: Miller, Louis H.

; APPLICANT: Peterson, David S.

; APPLICANT: Su, Xin-zhaun

; APPLICANT: Welles, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/487,826B

; FILING DATE: 10-SEP-1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Israelsen, Ned

; REGISTRATION NUMBER: 29,655

; REFERENCE/DOCKET NUMBER: NIH121.001CP1

; TELECOMMUNICATION INFORMATION:

; CITY: MADISON

Db 5555 AATATCATATAAATAATAAAATTAGTTAAAAATAAGTATTTTCATACAAAATACTAACT 5614

Qy 2929 taacataataattttaaatatcataattatacttttggtgaagtgcagctgggag 2988

Db 5615 TATAAGTATATCATATAATATTATATATATATATATTTTGTGTTTTTGATTGGGTGTAT 5674

Qy 2989 atacgtaaagtattttaaactatactttttgcaagcagttggtggtctaccaagagt 3048

Db 5675 ATAAGGCTATAAGTATATATGGGTTGCCATTATATATTTATATGTGAATAGATACATAT 5734

Qy 3049 gatcaaa-----gtttgagctgcctccaatgagccaaatttttgcccataatggataaa 3101

Db 5735 AAGTTAATATATTTATTTTGTGTATATGCTGTGTTAAGATAGATATGCATTACAGTTAAG 5794

Qy 3102 ggcaatttgttttagttcaactgctcacagaataatgttaaaatgaaattaaataagggtg 3161

Db 5795 GGTATAGATTTTTTTTTTTTTTTTTTTTGTACATATATATAAAAAATAGATAACTACAATA 5854

Qy 3162 gcctggtcacacacacaaaaaaaactaatgttggttggttgaattttattaccggaat 3221

Db 5855 TGCATATTACAAGAATAATATTGTATAAAATATATATATATATATATAAAGACAT 5914

Qy 3222 gtaattatttatattttaaataaaaattatgttatttagattcttaatttttggagcattc 3281

Db 5915 -TAAACTATACTAATAGGTAATTAGTTTTATTATATATCCTCTTTTATTATTATAATTTT 5973

Qy 3282 catactataatttcgtaacataattataaaaatatagtaataataaagtgaattaacctta 3341

Db 5974 TTTGTTTTACTTCTCTGCTCTTTTTTGTATTATAATATAACAATAATAAACATA 6033

Qy 3342 aattacaagcataatataaaattttgaatcaattaatttttattctatttttaatta 3401

Db 6034 TCAGTATTTGGAATATAAATAAATTATTCTACATATATGCATATATATATATATATA 6093

Qy 3402 atttagtctattttttcaaataaaaatttaaatctaaataaaaataatttt---cotta 3458

Db 6094 TAT 6153

Qy 3459 atgttgaacaactcatgtttatacttcaaaattataagttattattttaccotgatgatt 3518

Db 6154 ATTTTATATATATTTTAGTATATACTTTAAAGATATTATTAATATTATATAGTAGCATAT 6213

Qy 3519 tatttatttagtatattaattctgattataaattatggtgggatacaatcgctttccactaa 3578

Db 6214 ATGTATTATATATATAACAAATATTTTCATTATATAAATATATAGAACATGAACATTTT 6273

Qy 3579 atattttaactatgatt-tataaaattatttcaacatcgattatttacttattaatacat 3637

Db 6274 ATTAATAACTCATATTTGAATATATATATTATAATGTGATTTTACTTATTTTTTIAT 6333

Qy 3638 aatttatcataattttatggaattgagaccaagaacattaaagagaacaaattctataa 3697

Db 6334 ATTATACAATAAAATTTTGAATTCATAAAATGCATGAATACATAAAAAATACAACAA 6393

Qy 3698 caaagacaatttagaaaaaat-gtacttttaggttaattttaagtactcttaaccaaaaa 3756

Db 6394 AACAAATGATAAAACATTTTTTATTAATATAATAATATAATAATAATAATATTTTTTC 6453

Qy 3757 caaaaattcaaatcaaatgaactaaataagataatataacatacgggaacatcttacttgt 3816

Db 6454 CTGTTATTATTATTCATTTTTTTTTTGTATATATATATATATATAATAAATTATA 6513

Qy 3817 aatcttacattccccataattttattatgaaaaataactcttatttactcgaactaaatgt 3876

Db 6514 ATATATA---ACAACAAAAATTAATAATAATATATACTACTTTTAAATATAATAACAA 6570

Qy 3877 tgtcacaaattattatctaataaagaaaaaacacttaattttataacattttttcatat 3936

Db 6571 TACAAGAATATGTATCTATACAATTATATATATATGAATATATAAATATGATAGATAA 6630

Qy 3937 atttgaaga 3946

Db 6631 TATAGATAGA 6640

Search completed: September 3, 2000, 03:05:49
Job time: 28295 sec

RESULT 11
T43361
ID T43361 standard; cDNA; 974 BP.
AC T43361;

RESULT 12
 T43362
 ID T43362 standard; cDNA; 645 BP.
 AC T43362;
 DT 11-MAR-1997 (first entry)
 DE Cotton FbLate 2-82A gene cDNA clone A11 (FbLate-2).
 KW FbLate; promoter; fibre; transgenic plant; cotton; ds.
 OS Gossypium hirsutum.
 PN WO9639021-A1.
 PD 12-DEC-1996.
 PF 06-JUN-1996; U09449.
 PR 06-JUN-1995; US-467504.
 PA (MONS) MONSANTO CO.
 PI John ME;
 DR WPI; 97-042726/04.
 PT Plant fibre-specific, developmentally regulated FbLate promoter -
 PT useful for producing transgenic plants, esp. cotton, with altered
 PT fibre properties
 PS Claim 8; Page 56-57; 79pp; English.
 CC cDNA clones A8 or FbLate-1 (T43361) and A11 or FbLate-2 (T43362)
 CC correspond to RNAs prevalent in late development of cotton
 CC fibers. They were isolated from a 23-day cotton fibre cDNA

Db 61 TATTAATAATTATTAATAAAATTTATGGACTTTTGGACTGTCTGACTAATTTTCAGAA 120
Qy 1866 ttttattttggttttgggttttggtaatttttagataatttttaaatattctgcat 1925
|||||
Db 121 TTTTATTGGTTTGGGTTTGTGTAGTATTTTAGATAAATTTTAAATATTCTGCAT 180
Qy 1926 aattttctgttatttgaaggatgttcgaattttttcaaaatgaaacgtttaaga 1985
|||||
Db 181 AATTTTCTGTATTGTAAAGGATGTCGAATTTTTCAAAATGAAACGTTAAGA 240
Qy 1986 attttactactgcaaatcagaataagtgaatttttttagaagattaaataagtt 2045
|||||
Db 241 ATTTTACTACTGCAAAATCAGAAATAGTGAATTTGTTTTAGAAAGATTAAATAGTT 300
Qy 2046 agtattacgatttttagttttagtttgggaagtaattgttttgaacataattat 2105
|||||
Db 301 AGTATTACGATTTTAGTTTGTATTGGTGGAAAGTATGTATGTTTGAACATAATTAT 360
Qy 2106 ttgacaataaaggtttttagggaataaaccggaatatcttc-ttctttttgtaaa 2164
|||||
Db 361 TTGACAATAATTAAGTTTCTAGGAAATAACGGAATATCTCTTTTTTTTTGTAATA 420
Qy 2165 ttactaatgcaagaacaaacacgttttggggagcaataatctagctttaagtagtcag 2224
|||||
Db 421 TTACTAATGCAAGAACAAACACGTTTGGGAAGCAATAATCTAGCTTTAAGTAGTCAG 480
Qy 2225 tgtaactctcaaaatctggtcataactcttaggtgagtttggctgtctacagtagtaag 2284
|||||
Db 481 TGTAACTCTCAAAATCTGGTCATAACTCTAGGTGAGTTTGTCTGTCTACAGTAGTAAG 540
Qy 2285 tctatagaacttacctgacaaaacgacatgacgtcagggtcgaaatctacaacttttct 2344
|||||
Db 541 TCTATAGAACTTACCTGACAAAACGACATGACGTGAGGTCGAATCTACAACCTTTTCT 600
Qy 2345 ttttcttcaataacatggttgattcaagttccgatctataaattattacgattt 2404
|||||
Db 601 TTTTCTCAATTAACATATGTTGATTCAAGTTCCGATCTATAATAATTTATTACGATT 660
Qy 2405 atcaatttcaattaccttatatcatcctattataaataaagtcagttcaattcagttt 2464
|||||
Db 661 ATCAATTTCAATTACCTTATATCATCTATTATAAATAAAGTCAGTTCAATTCAGTTT 720
Qy 2465 cgaaagttcccaaaaatttgaattttataaatttttccctaaaacgaaatgattat 2524
|||||
Db 721 CGAAAGTTCCTTAAATTTTGAATTTTATTAATTTATCCCTAAAACCGAAATAGTGAT 780
Qy 2525 atctttcaaatttaagtttcatcttttcaatccgatttcaatttcaccttttataactct 2584
|||||
Db 781 ATCTTTCAATTTAAGTTTCATTTTCAATCCGATTTCATTTTCATCTCTTTATAACTCT 840
Qy 2585 ctattatctataattacataaatttcaaatattttgaattttacacttttagtccct 2644
|||||
Db 841 CTATGATCTATAATTACATAAATTTCAAACTAATTTTGAATATATACACTTTAGTCCCT 900
Qy 2645 aagttcaaaactataaattttcacttttagaattaatcattttcacatctaaagcatcaa 2704
|||||
Db 901 AAGTTCAAACTATAAATTTTCACTTTAGAAATTAATCAATTTTCACATCTAAGCATCAA 960
Qy 2705 atttaaccaaatgacacaaatttcatgattagttagatcaagcttttgagtcttcaaac 2764
|||||
Db 961 ATTTAACCAATGACACAAATTTCACTTTAGAAATTAATCAATTTTCACATCTAAGCATCAA 1020
Qy 2765 ataaaaatt-----acaaaaaaacaaacttaaatcatttatcaatttgaaacacaa 2820
|||||
Db 1021 CATAAAAATTACAAAAAATAAAACAACTTAAATCATTTTCAATTTGAAACACAAA 1080
Qy 2821 gcttgccgaatgctaagagcttaaaatggctcttttgggttttgggttgcacacgg 2880
|||||
Db 1081 GCTTGCCGGAATGCTAAGAGCTTAAAAATGGCTCTCTTTTGGTCTTTTGGTGGCAACGG 1140
Qy 2881 tggagagaagagggaatgaagattgaccatattttttattatgttttaacataata 2940
|||||
Db 1141 TGGAGAGAAGAGGGAATGAGATTGACCATATTTTTTATTATGTTTTAACATATAATA 1200

Qy 2941 ttaataatttaatacattatacttttggtaagtgtgacagtggtggagatacgtaaagta 3000
|||||
Db 1201 TTAATAATTTAATCATATTTATATCTTTGGTGAATGTGACAGTGGGGAGATACGTAAGTA 1260
Qy 3001 ttttaacattatacttttgcagcagttggctggtctaccgaagtgatcaaagtttg 3060
|||||
Db 1261 -TATAACATTATACCTTTTGAAGCAGTTGGCTGGTCTATCCAGAGTGATCAAAGTTTG 1319
Qy 3061 agctgccttcaatgagccaatttttggccataatggataaaggcaattgttttagttcaa 3120
|||||
Db 1320 AGCTGCCCTCAATGAGCCAATTTTGGCCATAATGGATAAAGGCAATTTGTTTAGTTCAA 1379
Qy 3121 ctgctcacagaataatgttaaaatgaaataaagtggtgctgctcacacacac--- 3177
|||||
Db 1380 CTGCTCACAGAATAATGTAAATGAAATTAATAAGTGGCTGGTGCACACACACA 1439
Qy 3178 aaaaaaaactaattgttgggttggtgaattttattattacggaatgtaattattatttta 3237
|||||
Db 1440 AAAAAAACTAATGTTGGTTGGTTGAATTTTATATTACGAATGTAATGTTATATTTTA 1499
Qy 3238 aaataaaattatgttatttagattcttaatttttggagcattccatactataatttctg 3297
|||||
Db 1500 AAATAAAATTATGTTATTAGATTCTTAATATTTT-GAGCATTCCTACTATAATCTCGT 1558
Qy 3298 a-acataatattaaaaatagtaataaagtgtaattaaactttaattcaagcataat 3356
|||||
Db 1559 ATACATAATATTAATAATAGTAATATAAAGTGAATTAACCTTAATTAACAGCATAAT 1618
Qy 3357 attaaatttgaatcaattaattttttattctattatttttaattatttagtctattttt 3416
|||||
Db 1619 ATTAATTTTGAATCAATTAATTTTATTCTATTATTTTAATTAATTAGTCTATTTTT 1678
Qy 3417 tcaaaaaaatttaattctaaataaaaaatttttcttaattgttgaacaactcatg 3476
|||||
Db 1679 TCAAAATAAAATTTAAATCTAAATAAAATAATTTTCTCTAATATT----- 1725
Qy 3477 ttactctcaaaattataagttattatttacccttgatgattttatttagtatataa 3536
|||||
Db 1726 -----ATTAATAATTTATTTCACATCATATTTACTTATTATAACATAAA 1773
Qy 3537 tctgattataattatggtgggatacaatcgctttccactaaatttttaactatgattt 3596
|||
Db 1774 TTAT----- 1777
Qy 3597 ataaatttttcaacatcgatatatttacttattataacataattttatcataattttatg 3656
|||||
Db 1778 -----AATAATTTATCATATTTTATG 1799
Qy 3657 gaatttgagaccaagaacacattagagaacaaattctataacaaagacaatttagaaaaa 3716
|||||
Db 1800 GAATTTGAGACCAAGAACATTAGAGAACAAATTCATATAACAAAGCAATTTAG-TAAA 1858
Qy 3717 aatgtacttttaggtaatttttaagtactcttaacaaacacaaaaattcaaatcaaatga 3776
|||||
Db 1859 AATGTACTTTTAGGTAATTTTAACTACTCTTAACCAACACAAAAATTCAAATCAATGA 1918
Qy 3777 actaaataagataataacatacaggaacattacttgtaattcttaccataaatt 3836
|||||
Db 1919 ACCAAATAAGATAATATAACATACAGAAATCTACTGTATTCTTACATTCCCGTAATC 1978
Qy 3837 ttattatgaaaaaatcttattactcgaactaaatgtgtcacaaattattatctaa 3896
|||||
Db 1979 ATATTATGAAAAGTAATATTATTACTGTAGCCAAATGCTCTCACAAATATTATCCAA 2038
Qy 3897 ataaagaaaaac--acttaattttataacatttttccatattttgaagattatattt 3954
|||||
Db 2039 AAAAAAATGTTGAATATAATTTTATAACATTTTTCATATATTGCAAGATTATATT 2098
Qy 3955 tgtatatttacctgaaaaattttgacatagattgagcaccttcttaacataatcccacca 4014
|||||
Db 2099 TGTATATTACGTAAAAATATTGACATAGATTGAACACCTCTTAAACATAATCCACCA 2158

|||||
Db 1973 TAAGAAATTTTACTACTGCAAAATTCAGAAATAGTGAATTTGTTTTTGAAGATTAAAT 2032
Qy 2041 aagttagtattacgatttttagtttgatttgggaaagtaagtgtattttgaacata 2100
Db 2033 AAGTTAGTATTACGATTTTGTGTTGATTGGTGGAAAGTAATGTATGTTTTGAACATA 2092
Qy 2101 attatttgacaataaagttttctagggaaataaacggaaatattctctctttttgt 2160
Db 2093 ATTATTGACAATAATTAAGTTTCTAGGGAATAACGGAATATCTTCTTTTGT 2152
Qy 2161 aaaattactaatgcaagaacaaacacgttttggggagcaaatatctagcttaagtag 2220
Db 2153 AAAATTACTAATGCAAGAACAAACACGTTTGGGGAGCAATATCTAGCTTAACTAG 2212
Qy 2221 tcagtgaactctcaaaatctggctataactctaggtgagtttctgtgtctacagtag 2280
Db 2213 TCAGTGAACCTCTCAAAATCTGGCTAATCTTAGGCTGAGTTCTGTGCTACAGTAG 2272
Qy 2281 taagctatagaaacttacctgacaaaacacatgacgtcaggtcgaatctacaactt 2340
Db 2273 TAAGCTATAGAACTTACCTGACAAAACGACATGACGTGAGGTGCAATCTACAACCTT 2332
Qy 2341 tctttttcttcaattaacatagttgattcaagttccgactataataattattacg 2400
Db 2333 TCCTTTTCTTCAATTAACATATGTTGATTCAAGTCCGATCTATAAATTTATTACG 2392
Qy 2401 atttatcaatttcaattacottatatcatcctattataaataagtcagttcaattcag 2460
Db 2393 ATTTATCAATTTCAATTACCTTATATCATCTTATTATAAATAAGTCAGTCAATTCAG 2452
Qy 2461 ttttcgaaagttcccaaaattttgaattttatataaattttccctaaacggaatag 2520
Db 2453 TTTTCGAAAGTCCCAAAATTTGAATTTTATTAATTTATCCCTAAACCGAAATAG 2512
Qy 2521 ttatatctttcaaaatttaagtttcttttcaatccgatttcaatttcatcctttataa 2580
Db 2513 TTATATCTTCAAAATTAAGTTTCAATTTTCAATCCGATTCAATTCATCTTTTATAA 2572
Qy 2581 ctctctattattataattacataaaatttcaaatattttgaaattttacactttagt 2640
Db 2573 CTCTCTATTATCTATAATTACATAAATTTCAAAATTAATTTGAAATATTACACTTAGT 2632
Qy 2641 ccctaagttcaaaactataaaattttcacttttagaatttaattttttcacatctaagca 2700
Db 2633 CCTAAGTTCAAACTATAAATTTTCACTTTAGAAATTAATCATTTTCACTCTAAGCA 2692
Qy 2701 tcaaaatttaacaaatgacacaaatttcatgattagtagcaagcttttgagcttca 2760
Db 2693 TCAAAATTAACCAATGACACAAATTTCACTTTAGAAATTAATCATTTTCACTCTAAGCA 2752
Qy 2761 aaacataaaaattacaaaaaaacaaacttaaaatcatttatcaatttgaacaacaaa 2820
Db 2753 AAACATAAAAATTACAAAAAAACAAACTTAAATCATTTATCAATTTGAACAACAAA 2812
Qy 2821 gcttggccgaatgctaagagcttaaaatggctcttctttgtttcttttgggtgcaaacgg 2880
Db 2813 GCTTGGCCGAATGCTAAGAGCTTAAATAGGCTCTTTTGTCTTTTGTGCAACGG 2872
Qy 2881 tggagagaagagggaatgaagattgacctattttttattattgttttaacataataa 2940
Db 2873 TGGAGAGAAGAGGGAATGAAGATTGACCATATTTTTTATTATGTTTTAACATATAATA 2932
Qy 2941 ttaataatttaatacattatacttttggatgtgacagtgaggagatcgtaaagta 3000
Db 2933 TTAATAATTTAATCATAATTATCTTTGGTGAATGTGACAGTGGGAGATACGTAAGTA 2992
Qy 3001 ttttaacattatactttttgcaagcagttggctgtctacccaagagtgatcaaagtgtg 3060
Db 2993 TTTTAACATTATCTTTTGAAGCAGTTGGCTGGTCTACCCAAGAGTGATCAAAGTTG 3052
Qy 3061 agctgccttcaatgagcaatttttggccataatggataaaggcaattgttttagttcaa 3120
|||||

Db 3053 AGCTGCCTTCAATGAGCCAATTTTGGCCATAATGGATAAAGGCAATTTGTTTGTTCAA 3112
Qy 3121 ctgctcacagaataatgttaaaatgaattaaaaaagggtggcctgcacacacacaaa 3180
Db 3113 CTGCTCACAGAATAATGTAAAAATGAAATTAAGTTGGCTGGTCACACACACAAA 3172
Qy 3181 aaaaaactaatgttggttggtgaattttattattacggaatgtaattatttttttttt 3240
Db 3173 AAAAACTAATGTGGTGGTGAATTTTATATTACGGAATGAATATTATTTTAAAA 3232
Qy 3241 taaaattatgttatttagattcttaatttttggagcattccataactataatttcgtaac 3300
Db 3233 TAAAAATATGTTATTAGATTCTTAATATTTTGGAGCATCCACTATAAATTCGTAAC 3292
Qy 3301 ataattataaaatatagtaataaaagtgaatttaactttaattacaagcataatatta 3360
Db 3293 ATAAATTTAAATATAGTAATATAAAGTGAATTAACCTTAAATTAACAGCATATAATA 3352
Qy 3361 aattttgaatcaattaattttttttctattatttttaattatttagtctattttttcaa 3420
Db 3353 AATTTTGAATCAATTAATTTTATTCTATTATTTAATTAATTTAGTCTATTTTTCAA 3412
Qy 3421 aataaaatttaaaatctaaataaaataaatttttcttaattgtgaacaactcatgttat 3480
Db 3413 AATAAAATTTAAATCTAAATAAAATAAATTTTCCCTAAATGTTGAACAACATCATGTAT 3472
Qy 3481 acttcaaaattataagttattatttaccctgatgattatttttagtattataattct 3540
Db 3473 ACTTCAAAATATAAGTATTATATTACCTTGATGATTATTATTAGTATTATAATTCT 3532
Qy 3541 gattataattatggtgggacataacgccttccactaaatttttaactatgattataa 3600
Db 3533 GATTATAATTATGTTGGGATACAATCGCTTTCCACTAAATTTTAACTATGATTATAA 3592
Qy 3601 atttatttcaacatcgatatatttacttattataacataattttatcataatttttgaaa 3660
Db 3593 ATTTATTCAACATCGTATATTACTTATTATACATAATTTATCATAATTTATGGAAA 3652
Qy 3661 ttgagaccaagaacattaagagaacaaattctataacaaagacaatttagaaaaaatg 3720
Db 3653 TTGAGACCAAGAACAATTAAGAGAACAAATCTATAACAAAGACAATTTAGAAAAAATG 3712
Qy 3721 tacttttaggtaatttttaagtactottaacaaacacaaaaattcaaatcaaatgaacta 3780
Db 3713 TACTTTTAGGTAATTTTAACTACTCTTAACCAACACAAAAATCAAAATCAATGAACATA 3772
Qy 3781 aataagataataacatacgggaacatcttacttgtaattctacattcccataattttat 3840
Db 3773 AATAAGATAATATAACATAGGGAACATCTTACTTGAATCTTACATCCCATATTTTAT 3832
Qy 3841 tatgaaaaaataattattactcgaactaaatgtgtgcacaaattattatctaaataa 3900
Db 3833 TATGAAAAATATCTTATATTACTCGAACTAAATGTTGCACAAATTATTATCTAAATAA 3892
Qy 3901 agaaaaacacttaatttttataacattttttcataattttgaaagattatatttggata 3960
Db 3893 AGAAAAACACTTAATTTTATAACATTTTTCATATATTGAAAGATTATATTGTGATA 3952
Qy 3961 ttacgtaaaaattttgacatagattgagcaccttcttaacataatcccaccataagtc 4020
Db 3953 TTTACGTAAAAATTTTGACATAGATTGAGCACCTTCTTAACATAATCCCACCATAAGTC 4012
Qy 4021 aagtatgtatgagaaattgtacaacaacgtggggccaaatcccacaaacacatctc 4080
Db 4013 AAGTATGTAGATGAGAAATGGTACAACAACGTGGGGCCAAATCCCACCAACCATCTC 4072
Qy 4081 tcattctctctataaaaaggcttgctacacatagacacaaatccacacacaaatcacgt 4140
Db 4073 TCATTCTCTCTATAAAAGGCTTGCTACACATAGACAACAATCCACACACAAATACAGT 4132
Qy 4141 tctttttcttatttattgataaccatggctcatagcattgtcaccctttcttcttttt 4200
Db 4133 TCTTTCTTCTATTATTAACCATGGCTCATAGCATTCGTCACCCCTTCTCTCTTTTC 4192

Db 3841 TATGAAAAAATCTTATATTACTCGAACTAAATGTTGCACAAATATTATCTAAATAA 3900
Qy 3901 agaaaaaaccttaattttataacattttttcatatattgaaagattatattttgtata 3960
|||||
Db 3901 AGAAAAACACTTAATTTTATAACATTTTTCATATATTGAAAGATTATATTTTGATA 3960
Qy 3961 tttacgtaaaaatattgacatagattgagacaccttcttaacataatcccaccataagtc 4020
|||||
Db 3961 TTTACGTAAAAATATTGACATAGATTGACACCTTCTTAACATAATCCACCATAAGTC 4020
Qy 4021 aagtatgtagatgagaaattggtacaacaacgtggggccaaatcccaccaaacatctc 4080
|||||
Db 4021 AAGTATGTAGATGAGAAATGGTACAAACACGCTGGGGCCAAATCCCAACCAACATCTC 4080
Qy 4081 tcattctctctataaaaggcttgctacacatagacaacaatccacacacaaatcacgt 4140
|||||
Db 4081 TCATTCTCTCTATAAAAGGCTTGCTACACATAGACAACATCCACACACAAATACAGCT 4140
Qy 4141 tcttttctctatttattgattaacatggctcatagcattcgccaccttttctcttttc 4200
|||||
Db 4141 TCTTTCTTCTTATTGATTAAACATGGCTCATAGCATTCGTACCCCTTCTTCTTCTTC 4200
Qy 4201 caacttttactcataagtgctcactagtacgggtgagccacactgtttcggcagcggt 4260
|||||
Db 4201 CAACCTTTACTCATAGTGTCTCAGTGTGACGGTAGCCAGCTGTTTCGGCAGCGGT 4260
Qy 4261 cgacgtttattcgagacacaaagcaacctcatcagagctcccaaatgggtctcaaaatc 4320
|||||
Db 4261 CGACGTTTATTCGAGACACAAGCAACCTCATCAGAGCTCCACAAATGGGTTCAAAATAC 4320
Qy 4321 gaaagcagcagagtgctgaatacgaagccagaaatacaaacagccaaagtatcacgaaga 4380
|||||
Db 4321 GAAAGCAGCAGAGTCTGAATACGAAAGCCAGAATACAAACAGCCAAAGTATCAGGAAGA 4380
Qy 4381 gtactcaaaacttgagaagcctgaaatgcaaaaggaggaaaaaaccctgcaaaaca 4440
|||||
Db 4381 GTACTCAAACTTGAGAGCCTGAAATGCAAAAGAGGAAAAAACCCTGCAAAACA 4440
Qy 4441 gcatgaagagtaccacgagtcacacgaatcaaaggagcaaaagagtacgagaagaaaa 4500
|||||
Db 4441 GCATGAGAGTACCACGAGTCACAGAAATCAAAGAGCAAAAGAGTACGAGAAAGAAA 4500
Qy 4501 tctgcagcaattccccggggcgctgcagcggtagcgaagattctcgggcccgctgcagcctt 4560
|||||
Db 4501 TCTCGACGAATCCCCCGGGCGCTCGACGGCTAGCGAAGATCTTCGGGCCCGCTCGAGCCTT 4560
Qy 4561 gaatcatatgacactggtgcatgtgccatcatcagtaatttcatggtatattcgtaa 4620
|||||
Db 4561 GAATCATATGACACTGGTGCATGTGCCATCATATGCAGTAATTTATGTTATATCGTAA 4620
Qy 4621 tatatagttataaaaaagatggtgattgggaatgtgtgtgcatctccatgcact 4680
|||||
Db 4621 TATATAGTTAATAAAAAAGATGGTGATTGGGAATGTGTGTGTCATTCTCCATGCAC 4680
Qy 4681 aatggtgaatctctttgcatacatagaaattctaatggttatagtttatgttagtgt 4740
|||||
Db 4681 AATGGTGAATCTCTTGCATACATAGAAATCTAAATGGTTATAGTTTATAGTTATAGTGT 4740
Qy 4741 atgttgtagtgaaattatttaattgtgtatctaatgttaacatcacttggttgatt 4800
|||||
Db 4741 ATGTTGTAGTGAAATTAATTTAAATGTTGATCTAATGTTAATCATCATTGGCTTGATT 4800
Qy 4801 tatgttatgttatgtattttactttaatgatattgcatgtattgttaatttaacattgct 4860
|||||
Db 4801 TATGTTATGTTATGTTATTTACTTTAATGATATTGCATGATTGTTAATTAACTATGCT 4860
Qy 4861 tgatcattatactcttctactattaattataaatggcactgttttgttaaaactttttac 4920
|||||
Db 4861 TGATCATTATACCTCTTACTATTAAATATAAATGGCACTGTTTGTGTTAACTTTTAC 4920
Qy 4921 aagtttaagacatgtataatatatgacaatatataattcaggttttagttcaatgttagct 4980
|||||
Db 4921 AAGTTAAGACATGTATAATATATGACAAATATAATACAGGTTTTAGTTCATGTAGCT 4980

Qy 4981 atcttagtagtattgtatgatcttaattacatttaacaaattccacttaaaattttta 5040
|||||
Db 4981 ATCTTAGTATGTTATTGATGATCTTAATTACATTAAACAAATCCACTTAAATTTTAA 5040
Qy 5041 taaataatacaaaataattattgtaataataacattaaatgcaacaaaaatgaaataa 5100
|||||
Db 5041 TAAATAATAACAAATAATTATTGTAATATAATACATTAAATGCAACAAAAATGAAATAA 5100
Qy 5101 ataaaaataaaatagcaaaatattgttataatattgtaataataatgtaccatattctta 5160
|||||
Db 5101 ATAAAAATAAATAGCAATAATGTTATAATATTGTAATATAATGTACCATATTCTTA 5160
Qy 5161 actgaaatagggtctaacctataatccctaaatttcagtttaaatattttttatctac 5220
|||||
Db 5161 ACTGAAATAGGGCTAACCTATAATCCCTAAATTTTCAGTTTAAATATTTTATACCTAC 5220
Qy 5221 catattattagaactctttttaaatatttaaaatttaattataccaatttaattaaac 5280
|||||
Db 5221 CATATTATTAGAACCTTTTTAAATATATAAAATTTTAAATATACCAATTAAATAAAC 5280
Qy 5281 tattaattatcttaactaaaatctaaatttttatttaacctattaataaattcctaatta 5340
|||||
Db 5281 TATTAATTATCTTAACATAAACTAAAAATTTTATTAACCTATTAATAAATCCATAATTA 5340
Qy 5341 tcttatctaattttaaaactctaattatcctaatttaatttaattcctaattatctta 5400
|||||
Db 5341 TCTTATCTAATTTAAACTCTAATTATCTTAATTAAATTTAAATCTTAATTATCTTAAT 5400
Qy 5401 ttgtaacctctccaccagctagatgtggaccggaatccgggagattacatcgccat 5460
|||||
Db 5401 TTGTAACCTCTCCACCAGCTAGATGCTGGACCCGAATCCGGGAGATTACATCGGCCAT 5460
Qy 5461 tgagatggcgatcaggggttggtggtggtggtacccaattgcgcctatagttagttcgt 5520
|||||
Db 5461 TGAGATGGCGTGATCAGGGTTTGGCGCGCGGTACCAATTCGCCCTATAGTGAGTTCGT 5520
Qy 5521 attacggtgctcactgcgtccggttt 5547
|||||
Db 5521 ATTACGCGCGCTCACTGCGTCCGGTTT 5547

RESULT 2

T73870

ID T73870 standard; DNA; 5518 BP.

AC T73870;

DT 26-JAN-1998 (first entry)

DE Cotton fibre promoter clone 4-4(6) construct, pCGN5610 (Version II).

KW promoter; fibre-specific; transcriptional factor; promoter;

KW altered phenotype; colour; melanin; indigo; ss.

OS Gossypium hirsutum cv. coker 130.

PN W09640924-A2.

PD 19-DEC-1996.

PF 07-JUN-1996; U09897.

PR 07-JUN-1995; US-480178.

PR 01-JUL-1996; ZA-005572.

PA (CALJ) CALGENE INC.

PI McBride K, Pear JR, Perez-Grau L, Stalker DM;

DR WPI; 97-052325/05.

PT DNA construct contg. gene of interest controlled by cotton fibre

PT transcriptional factor - used to produce altered phenotype cotton

PT fibre cells expressing genes affecting pigmentation

PS Example 5; Fig 3A-J; 95pp; English.

CC The present sequence is a 4-4 cotton fibre expression cassette (version
CC II) from promoter construct pCGN5610. The lambda genomic phage clone used
CC to form this construct was designated 4-4(6). DNA constructs containing
CC cotton fibre-specific transcriptional factor promoters are useful to
CC produce cotton fibre cells with altered phenotype, especially altered
CC colour. Genes involved in the production of melanin (e.g. tyrosinase
CC gene and ORF438 encoded protein from Streptomyces antibioticus) and
CC indigo (mono-oxygenase genes possibly in conjunction with a
CC tryptophanase gene) are of interest. The promoters of the invention are
CC reliable and permit expression of a protein selectively in cotton fibre

Qy	1	actaaagggaacaaaagctggagctccaccgcggtggcgccgctctagaactagtggat	60
Db	1	ACTAAAGGGAACAAAAGCTGGAGCTCCACCgcggtggcgccgctctAGAActAGTGGAT	60
Qy	61	ccccctggactaacaacaaacatgggaagatttgcgtgtaaaaaataaaagaagcttact	120
Db	61	CCCCCTGGACTAAACAAAACATGGGAAGATTTCGTGTAaaaaataaaAGAAGCTTACT	120
Qy	121	caataaacactttgtgaattgtatacaaaagactcaatgaaaaacaataactcaatacact	180
Db	121	CAATAACACTTTGTGAATTGTATCAAAAAGACTCAATGAAAAACAATAACTCAATACACT	180
Qy	181	tttttctactgatttacatcctttatataggtgaaactacaacaacttttagctaaaaaa	240
Db	181	TTTTTCTACTGATTACATCCTTTATATAGCTGAAACTACAACTTTAGCTAAAAAA	240
Qy	241	ataggataacctaatagcaaaatcacatcagatattaaccatgattttagctaaccat	300
Db	241	ATAGGATAACCTAATAGCAAAATCACAATCAGATATTAACCATGATTTTAGCTAACCAT	300
Qy	301	ttaacaactttattgaaactaatttgaattttcatctgctgatatgcccaagattttag	360
Db	301	TTAACAACCTTTATTGAAACTAATTTGAATTTTCATCTGCTGATATGCCCAAGATTTTAG	360
Qy	361	gccactaaccgatttgggtggaactttaacatgctatgcaatttgtaactgtttgaaaca	420
Db	361	GCCACTAACCGATTGGGTGGGAACTTAACATGTCATGCATTTTGTAAGCTTTTGAAACA	420
Qy	421	agttttttgcattattttactatagaactgtttgattaggttgagttacacactgagct	480
Db	421	AGTTTTTTGCATTATTTTACTATATGAACCTGTTTGATTAGGTGAGTTACACACTGAGCT	480
Qy	481	tgtaagctcactcaaaattttctaatttctaaggtgatcagcaaataggaccggggcgg	540
Db	481	TGTAAGCTCACTCAAATTTTCTAATTTCTAAGGTGATCAGCAAACTTAGGACGGGCGG	540
Qy	541	cgtacgagagctcggattgttttctagttaataaataagacgatttatgtttttaaact	600
Db	541	CGTACGAGAGCTCGGATTGATTTTCTAGTTAATAAATAAGACGATTTATGTTTTTAAACT	600

Qy 601 attatggaccttttggactatgtaacgttttgggaccttatttttggttttttatctgtc 660
| | | | |
Db 601 ATTATGGACTTTTGGGACTATGTAACGTTTTGGGACTTATTTTTGTTTTATTTCGCT 660

Qy 661 tttttggatttagtaattattatttttaaactgcaaaattatatgttttacaaactaag 720
| | | | |
Db 661 TTTTGGATTTAGTAATTATTATTTTAACTGCAAAATTATATGTTTTACAACATAAG 720

Qy 721 tcacagttttcaaaattccataacttagaatttttcgctgcaaaataagaatcatctta 780
| | | | |
Db 721 TCACAGTTTCAAAATCCATAAECTAGAATTTTCGCTGCAAAATAAGTAATCATTTA 780

Qy 781 agtgtttttctgtataaaaaataaaataattttaacgagtatcttctocaaaaattgg 840
| | | | |
Db 781 AGTGTTTTTCTGTATAAAAAATAAATAAATTTAACGAGTATTTCTCAAAAATTGG 840

Qy 841 aaattgattaccaaaatttagtatgtcaaacacatgtttatatgttacagggcgatatac 900
| | | | |
Db 841 AAATGATTACC AAAATTAGTATGTCAAAACACATGTTATATGTTACAGGGCGATATC 900

Qy 901 gctcaggcaaatcaacatctagcggggtttggagtggttacagggcgagtgggctcatttt 960
| | | | |
Db 901 GTCAGGCCAAATAACATCTAGCGGGGTTTGAGTGTTACAGGGCGAGTGGGCTCATTTT 960

Qy 961 gagtaagtatatgttagggccgagttttagattgcataattcaaggtcaaagattttgtaaa 1020
| | | | |
Db 961 GAGTAAGTATAGTTAGGCCGAGTTT TAGATTGCATATTCAAGGTCAAAGATTTGTAAA 1020

Qy 1021 cttcgatgaatgatatagtatgattgccgattaacgaaatatgtttttcttttgtgtg 1080
| | | | |
Db 1021 CTTCGATGAATGATATGATGATTGCCGATTACGAAATATGTTTTTCTTTTGTGTG 1080

Qy 1081 tgttttatctcgtgtgataagtatatagtagttttattccaattcttatggcatgtgac 1140
| | | | |
Db 1081 TGTTTTATCTCGTGTGATAAGTATATAGTATGTTTTATTCCAATCTTATGGCATGTGAC 1140

Qy 1141 atttgggtatttctaataaattgattgtttatttgaaatctgatgcattctgttctac 1200
| | | | |
Db 1141 ATTTGGGTATTCTAATAAATTGATTGTATTATTGAAATCTGATGCATCTGTCTAC 1200

Qy 1201 aaagcatggaattctcatgcctactgctttctgttaaagatcacgattgcaagtttaacatg 1260
| | | | |
Db 1201 AAAGCATGGAATCTCATGCCTACTGCTTCTGTGTTAAAGATCACGATTGCAAGTTAACATG 1260

Qy 1261 ctactatttttgattttgtccttgcatgctatgtcacattacatggggtgggatgatata 1320
| | | | |
Db 1261 CTACTATTTTGATTTTGTCTCTGCATGCTATGTCACATTACATGGGGTGGGATGATAT 1320

Qy 1321 ggtaaggaggaaagttttgacagtttaatgatttgcactatctggtgggttaaccacatat 1380
| | | | |
Db 1321 GGTAAGGAGGAAGTTTGACAGTTTAATGATTGACATATCTGTTGGTTTAACCACATAT 1380

Qy 1381 ttgttatggcatcttgactgcgggttatgggtgcgaccgccatatctgttctggaat 1440
| | | | |
Db 1381 TTGTTATGGCATCTGACTGCGGTTATGGTGGCTCGACCGCCCATATCTGTTCTGGAAT 1440

Qy 1441 ttactctgactctgggtgcatgtgtcacaattatttgggtgtgttttggatggacca 1500
| | | | |
Db 1441 TTACTCTGACTCTGGTGGCATGTGTCAAAATTATTGTTGGTGTGTTTGGATGGACGA 1500

Qy 1501 gtctgtgggaaactcattttggtgtgttgcggagttgggttaggaaattttcgaaaaaaatt 1560
| | | | |
Db 1501 GTCGTGGGAACCTATTGTTGTGTGTCGGAGTTGGGTAGGAAATTTTCGAAAAAAATT 1560

Qy 1561 tgcattgtgtttttctgaaaaattatgcattaacataatcatgcattctcaattttggtc 1620
| | | | |
Db 1561 TGCATTGTGTTTTTCTGAAAAATATTGCATTAACATAATCATGCATTCTCAATTTGGTC 1620

Qy 1621 aattgaacgtttataaaattctctatgatatacctgatctgtttattacattatatgtgtt 1680
| | | | |
Db 1621 AATTGAACGTTATAAAATCTCTATGATATCCGATCTGTTTATTACATTATATGTTT 1680

Qy	3232	attttaaaaaaaattatgtttatttagattctttaattttggagcattccataactataa	3291
Db	17044	ATATATATATATATATATATATACAAATTAATTAATTCGATTTAGTGATTAAAAATAAATAT	16985
Qy	3292	tttcgtaacaataattataaaatataagtaataaagtgtaattaactttaaattacaagc	3351
Db	16984	TTTATTATACCTTATATAATTAAATTGAAATTAATATATATATATATATAAATATA	16925
Qy	3352	ataataataaattttgaatcaattaattttttattctattatttttaataatttagtcta	3411
Db	16924	TGAATTGAATTTTATAAAAAATCATTTTAAATTTTATTATATATAAAAAATATTTTATT	16865
Qy	3412	ttttttcaaaaataaaattttaacttaaaaaataatttttcttaagtgtgaacaac	3471
Db	16864	AATTATTAAAAATAATTTTATTATAAAAATTTATTAT-----AAAAAT	16819
Qy	3472	tcattgtataacttcaaaattataagttattattaccttgatgattattttattagtagt	3531
Db	16818	AGTTTATTAGGTATAAATTAATAAATCATTTTTTTTTAAAAAATAATTTTAAAGTT	16759
Qy	3532	attaattctgattataaattatgggtgggatacaatcgcttccactaaatattttaactat	3591
Db	16758	TTAATTATACAATAAATTTATGAATAGGGGGAATAAATTTATTTTCATTTTTTATATAT	16699
Qy	3592	-gatttataaaatttatttcaacatcgtataatttacttattataacataaatttatcataat	3650
Db	16698	ATATATATATATATAAATTAATTTATTTTTCAGATTAGTGATTAAAAATAAATTTTATTAT	16639
Qy	3651	tttatggaaattgagaccagaacaacattaagagaacaaattctatacaaaagacaattta	3710
Db	16638	ATTTATATAAATTTA-----ATTGAAATTAATAATTTATATATATATATATATA	16590
Qy	3711	gaaaaaaattgtacttttaggttaattttaagtactcttaaccaaacacaaaaattcaaatc	3770
Db	16589	TAAATATAAATGAATTTTTTAAAAATATTTTTTAATTTTATTATATAAATAAATTTTC	16530
Qy	3771	aaatgaactaaataagataataataacatacggaacatcttacttgtaactctacattccc	3830
Db	16529	TTATTAAATTTTAAATAAATTTTATTATATAAATAAATTTATTATAAA---AATAGTTTA	16473
Qy	3831	ataattttattatgaaaaataatcttatattactcgaaactaaatgtgtcacaaattatt	3890
Db	16472	TTAAGTATAAATTTAATAAATAAATTTTTTTTTAAAAAATAATTTTTTAACTTTTAAT	16413
Qy	3891	atctaaaaaagaaaacacttaatttttataacatttttcatatatttgaagattat	3950
Db	16412	ATATAA-----TAAATTTATGAATAGGGGGAATAAATTTATTTTCATTTTACATATATA	16359
Qy	3951	attttgtatatattacgtaaaaaattttgacatagattgagcaccttcttaacataatccc	4010
Db	16358	TATATATATATATATATACAAATTAATTAATTCAGATTTAGTGATTAAAAATAAATTTT	16299
Qy	4011	accataagtcgaagtatgtagatgagaaattggtacaaacaacgtggggccaatcccacc	4070
Db	16298	ATTATACTTATATAAATTTAATTGAAATTAATTTATATGTATATATATAAATATATGA	16239
Qy	4071	aaaccatctctcattctctctctataaaaggcttgctacacatagacaacaatccacac	4130
Db	16238	ATTGAATTTTATAAAAAATCATTTTAAATTTTATTATATATAAATAATTTTATTAAAT	16179
Qy	4131	aaatacacgttcttttcttctatttgatttaaccatggctcatagca---tctgcaccc	4187
Db	16178	TATTTAAAAATAATTTTATTATAAATAAATTTATTATATAAATAAGTTTATTAAAGTATAAT	16119
Qy	4188	tttctctcttttccaacttttactcataagtgctcactagtgcgcgtagccactgt	4247
Db	16118	TTAATAAATCATTTTTTTTTAAAAAATAATTTTTTAAAGTTTAAATATACATAAAT	16059
Qy	4248	ttcggca----ggcgctcgagctttattcgagacacaaagcaacctcatcagagctcccac	4303
Db	16058	TTATGAATAGGGGGAATAAATTTATTTCATTTTTTTATATATATATATATATATAAT	15999

VERSION AL031746.9 GI:6594243
KEYWORDS HTG.
SOURCE malaria parasite *P. falciparum*.
ORGANISM *Plasmodium falciparum*
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 67970)
AUTHORS Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.
and Barrell, B.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1998) *P. falciparum* Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
COMMENT On Dec 16, 1999 this sequence version replaced gi:5763807.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
sequence is unfinished and does not necessarily represent the
correct sequence. Work on the sequence is in progress and the
release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from *E. coli*, yeast, vector,
phage etc.

FEATURES
source Location/Qualifiers
1. 67970
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="1"
gene complement(1748..3276)
/gene="MAL1P3.01"
CDS complement(join(1748..2598,2748..2848,2990..3276))
/gene="MAL1P3.01"
/note="MAL1P3.01, conserved hypothetical protein, len: 412
aa, similarity: UPF0006 family eg to
YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta
scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
overlap)"
/codon_start=1
/product="conserved hypothetical protein, UPF0006 family"
/protein_id="CAB63556.1"
/db_xref="GI:6594244"
/translation="MKLVFPHYIKYINVLVYISIIPLKSNLSKIYNDLRYISTVNYKYV
LQIKKRSNLKKNHNRKMDNESSFIDIGSNLTKMFDGVYNSKKHENDLQNVLNRAK
NNVVDKIIITCTCLAEIDKSLKICETIDPEGKFLYLAVHPTNCEYIDKNKHEEKE
IAKKEYEPIFYKNEQVNSKMGNGNKKICDGEKDMNNLNEILLENLDTIPGPKY
NEKDKYLENLKNNKIIPNRVIGCEIGLDFDRLYFCSKYIQIKYFIPQLKLVQMFN
LPMFLHMRNCSETFPKIVDIYKFLFEKNGGVIHSPTDKEDIVHIIQVQNKLYIGVG
CSLKSLENINAVKKIPLNLLLETDAPWCGVKRTHASIEYIKDYEKRAYTNLKKRIK
IICDDNTIPKERNEPYNIA"
misc_feature complement(2599..2610)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"
misc_feature complement(2742..2747)
/gene="MAL1P3.01"
/note="potential splice donor sequence, atg/gttaa"
misc_feature complement(2849..2861)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"
misc_feature complement(2984..2989)
/gene="MAL1P3.01"
/note="potential splice donor sequence, aaa/gtaaaa"
gene 5005..5496
/gene="MAL1P3.02"
CDS 5005..5496
/gene="MAL1P3.02"
/note="MAL1P3.02, hypothetical protein, len: 163 aa,
contains possible signal sequence"
/codon_start=1
/product="hypothetical protein, MAL1P3.02"
/protein_id="CAB63557.1"
/db_xref="GI:6594245"
/translation="MKLLNRRFVFLCPHIIILFFLNSVVLGNNRRNINFHETENAAK
AMRKLSSGEINSIKLDNGDELKIKLNDKHKDSTKWDKSYSPISNLEEEKYSQTLDFR
KKQEIENAWTKIEDRQEPYILNNDIEINIAIRFVLENNFDELYIQSFQSLIDIIQS

LNN"
misc_feature 8020..10389
/note="possible cen1, region of very high [A+T] content"
gene 14884..20352
/gene="MAL1P3.03"
CDS 14884..20352
/gene="MAL1P3.03"
/note="MAL1P3.03, putative ABC transporter, len: 1822 aa"
/codon_start=1
/product="putative ABC transporter"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/translation="MTTYKENVGISNKGKKNKSCQNSIFLNFSLFSDWIRPLINDLIK
GDIQELPNICRNFDPVYASKLEENLRDIEVEDSEFYSEKNSNEHVLHHCNSNDASE
KKVYNVYVNHILSKITFKFRILLIISFYILETLIVTLGGKPIDYMRILEGQKIPV
YISFLKDFKVFSGLVVVMIMFPHLFFALLHFFHFLFTINLKVSLMYFLYKINLCSNN
NHLQNPDAFYNTYRKFSQTEIDEISRDLFISIGNKASSSSGKNNKNNKNNKNNKNN
DYIINFIKSTKKMEKDSLNENRSLPNVNIYIMFSDVPSVTFVTSCINLNVFVKIF
MSFYVFIKIGSNSVGIAIWLISALYFELFELPSLFKSKYLIYRDKRIDNMHVL
KEFKLIKMPWESFAFKYINIFRMKMKYCKIRLYLSNIGVFISSISDIEVEVFFI
YKDLRLNKKEIEIKFTSIIMPLYVYKILISNVANFPLVNNVMEGIVNKRNNYINDH
LYINDIKNYFMYRTRNEDYNIYVDKTFIQENITSHDDGTSNHLKHLKLVKNNLTN
MFYFFFYHKMNYHKNIINQKILSGLLKNNVDNTNKKICFQEHKSNTYNTYNSHIE
KKEEYENIHSSNSTMSNEPKKKNNEYIKLENCSEGLSDYDNKCDNDHILKNINPN
LKRNLALIGNVGSGKSAFFHSILGDFMTHGNLYIENTFFKMPILYVQNSWLFMG
NIRSMILFGNEYNPLIYKTYILQSELLNDLSTIEHGMKYIINDDHNLKGGQKVRICLA
RALYEHYIEMHKLCTDYKELIQPNEILDKDLINNNKISSYNNKSKLVNYNIPFNEN
YLQKCLMDNNFYLYLLDDIFTSLDPSISKKIPSNLFCKEDNISFKDNCSFIISMNS
TLDNPLIEDILDNVQYEVNIFBEIQDKTLKRGNISEYMERNNLITKESHWGYSNLT
IDYTRIKLFEDEVNLHVHKNMKIYKAYFVKGTBSVSFEIDSINKEYIKMKKKNY
KKEHNNKNNKNNNNNNNNKDDHININMNDNHRNYNDINLGNSTDDSPVSSLGNE"
YTLDTYTSNNSDKEEIVKPKYDTHHEFNKSSMPFVSSSSNNINPNPFYEDNSS
FKGSISLEYLYWYQGVGFLLTSVIFMLISIFTDEIKFVFLTMMSIISKNNKESD
TILQKQVRYLEYFVILPIISLVTSGICFMSIYGNITSKIVHNNILYSILNAPLIIF
YNNLGNINIRFIDISAFDYGLFKRIYKAFIFFRICLSSLLIYMRDCIFIPFV
IILYIFVFKRFSRGCKEQRLYLSCHTPLCNIYSNALSNGKNIINIKNTYHLDVYE
HYNNFRISYFFKWLINWASLYIKIFILLLTYYIIMPHLYASGIKLYKEKNYVRI
LSTLYGICISFARLGVIIKFLCDYTHIEKEMCCVQRLKEFAKISNKENASMNKENEL
NVITTQTYKEKNENISDKISAIVEYKVNLSIINSQDDSKKKYGIKFENVYVSYK
KKIPLVNGTYKYIDEEPSLKNIMYALKNQKIGIVGKSGAGKSTILLSILGLINISQ
KITVEGRDIRTYNRKGEDSIIIGLAQSSFPVYNNIRTFIDPYNNTDDEIVHALKLN
GINLGKNDLYKYMHKQDKMSYKKIITQSKVINQSDNTILLTNDICRYLSYLVLYLN
RHXYKIIILDEIPFNLNNSVHDELNSFLICKAKSFNYIIRNHPNNTVLIISHANT
LSCDDYIYVLRKGEITRYCSYEDVKTQSELSHLEMD"
rRNA 23896..31533
/gene="rRNA"
/note="region containing small subunit, 5.8S and large
subunit rRNA genes and spacer regions"
gene 23896..31533
/gene="rRNA"
gene complement(31966..32775)
/gene="MAL1P3.04"
CDS complement(join(31966..32476,32675..32775))
/gene="MAL1P3.04"
/note="MAL1P3.04, conserved hypothetical membrane protein,
len: 203 aa, similarity: *P. falciparum* chromosome 2,
PFB0110W, O96126 predicted integral membrane protein (255
aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
in 191 aa overlap)"
/codon_start=1
/product="conserved hypothetical membrane protein,
MAL1P3.04"
/protein_id="CAB63559.1"
/db_xref="GI:6594247"
/translation="MKKSYTFINVTILLFLTLFLTYNYDYTSKTKFNNKIDIN
RFKRIIAEASEEQRYWEEDFLINNEELIRPEHNDSPYLPEHYENIDKINELINS
TKIWKETIKMRQNYEKETDMNNHNRDPMHMYKWNIIYLYVHKHINILKDLTNPI
HDKETITTWIKWQEDIEYFLNQLQVEWLRILTLFLYKNKE"
misc_feature complement(32477..32486)
/gene="MAL1P3.04"
/note="potential splice acceptor sequence"
misc_feature complement(32669..32674)
/gene="MAL1P3.04"

Qy 1810 aaataattatttaattaaaatttatggacttttggactgtctgactaattttcagaatttt 1869
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98635 AAATAAATCTTAATAAATAATTTTTTGATAGATTTCCTAGGATAATATGAATATTTC 98694

Qy 1870 attttggtttgggttttggtaatttttagataattattttaaattttctgcataatt 1929
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 98695 GTAAAAAATAATAATAAAATGATATTTAAATATTATATTAACT--CAATATTAAAT 98752

Qy 2321 agggtcgaatctacaactttctcttttcttccaattaacatattggtgattcaagttccg 2380
| | | | | | | | | | | | | | | | | | | | | |
Db 17899 ATTAATTATTATATATTATAAAATTATATATTATTGAATATTATTATATAATATATATATA 17958
Qy 2381 atctataataatttattacagtttatcaatttcaattacccttatcatcctattataaa 2440
| | | | | | | | | | | | | | | | | | | | | |
Db 17959 TATATAGAAAAAATAAATTATTTTAAATAAATTAAATATAAATTTTTAAAAATTTCTTAAA 18018
Qy 2441 tataagtcagttcaattcagttttcgaaagttcccaaaaatttgaattttataaattt 2500
| | | | | | | | | | | | | | | | | | | | | |
Db 18019 TGTATTATTTTTATAAAAAATATTATATAATAAAATCATGTTTTTTAAAAATAAACAA 18078
Qy 2501 attccctaaaaccgaaatagttatatcttctcaaaatttaagtttcatttttcaatccgatt 2560
| | | | | | | | | | | | | | | | | | | | | |
Db 18079 AAAATTTTTAATAATAAATTTTATAATGAATATAAATTATTTATTTTCAATTTTTTT 18138
Qy 2561 tcaatttcacatcttttataactctctattatctataattacataaatttcaaaattattt 2620
| | | | | | | | | | | | | | | | | | | | | |
Db 18139 AAAAAATTTTTAAAAAATAAATTTTTTTTTAAAAAATACTATATACTAATTATATAA 18198
Qy 2621 tgaatttttataacttttagtcctcaagttcaaaactataaaatttcaatttagaataaa 2680
| | | | | | | | | | | | | | | | | | | | | |
Db 18199 TTAATAGATATTATATATATATAAATATTTAATATATTATTATATATCTAATAATTAA 18258
Qy 2681 tcatttttcacatctaagcatcaaaatttaaccaaatgacacaaatttcattgattagtag 2740
| | | | | | | | | | | | | | | | | | | | | |
Db 18259 ATAAAAATTTTAAATTTAAAAATGTAGATATAAATTATAAAAAATTATATTCTCATAT 18318
Qy 2741 atcaagcttttgagctctcaaaacataaaaattacaaaaaaacaaactaaaatcat 2800
| | | | | | | | | | | | | | | | | | | | | |
Db 18319 TTATTATTATTAAATTTAATTTATATAAATAATATATGATTAAATTAATTATTATATAT 18378
Qy 2801 ttatcaatttgaacaacaagcttggcgaatgctaagagcttaaaaatggctctctttg 2860
| | | | | | | | | | | | | | | | | | | | | |
Db 18379 TTATAAATTTATATATTATTGAATATTATATAATATATATATATATAGAAAAATTAA 18438
Qy 2861 tttctttttgttgcaaacgggtggagagaagagggaaatgaagattgaccatatttttta 2920
| | | | | | | | | | | | | | | | | | | | | |
Db 18439 ATTATTAAATAAATTTAATAATAAATTTTTTAAAAATTTCTTAAATGTATTATTTTATAA 18498
Qy 2921 ttatgttttaacatataataattaataattt---aatcataattatacttttgggaattg 2976
| | | | | | | | | | | | | | | | | | | | | |
Db 18499 AAAATATTATATATAATAAATCATGTTTTTTAAAAATAAACAAAAAATTTTAAATAAT 18558
Qy 2977 gacagtggtggagatacgtaaagtatttcaactataacttttggcaagcagttggctggt 3036
| | | | | | | | | | | | | | | | | | | | | |
Db 18559 AAATTTTATAATGAATAATAATTATTTATTTTCAATTTTTTAAAAAATTTTAAAAA 18618
Qy 3037 ctaccaagagtgatcaaaagtttgagctgcttcaatgagccaatttttggccataatgg 3096
| | | | | | | | | | | | | | | | | | | | | |
Db 18619 AAAATAATTTTTTTTTTAAAAAATACTATATACTAATTATAAATTAATAGATATTATA 18678
Qy 3097 ataaaggcaatttgttagttcaactgtctcacagaaataatgttaaataagaattaaataa 3156
| | | | | | | | | | | | | | | | | | | | | |
Db 18679 TATATATAAATATTTAATATATTATTATATCTAATAAATTTAAATAAAAAATTTTAAAA 18738
Qy 3157 aggtggcctgggtcacacacacaaaaaaactaattgttggttggtgaattttattattac 3216
| | | | | | | | | | | | | | | | | | | | | |
Db 18739 TTTAAAAATGTAGATATAAATTTATAAAAAATTATATTCTCATATTATTTATTATTAAAT 18798
Qy 3217 ggaatgtaattattattttttaaataaaaattatgttatttagattcttaattatttggag 3276
| | | | | | | | | | | | | | | | | | | | | |
Db 18799 TAATTTATATAAATAATAATAATGATTTAATTAATTATTATATATTATAAATTTATATAT 18858
Qy 3277 cattccatactataatttcgtacataataattaaaaatagtaataata----- 3324
| | | | | | | | | | | | | | | | | | | | | |
Db 18859 TATTGAATATTATATAATATATATATATATATAGAAAAATAAATTAATTATTTAAATAATT 18918
Qy 3325 ----aagtgttaataacttttaaattacaagcataatattaaatttgaatcaattaatt 3379
| | | | | | | | | | | | | | | | | | | | | |
Db 18919 AATATAAATTTTTTAAAAATTTCTTAAATGTATATTTTTATAAAAAATATTATATAAT 18978
Qy 3380 tttatttctattatttttaatttaattttagtctatttttcaaaataaaatttaaatctaaa 3439

Qy 3440 taaaanaaatTTTTccttaagtgttgaaacaactcatgttatacttcaaattataagtat 3499
|| || | ||| | || || | || || | |||| | ||
Db 4123 TATAATTATTATTTTTCATTTTTTTAAAAAAAATTTTTTAAAAAAAATATTTTTTTT 4182

Qy 3500 tatattacccttgatgattttatttagtatattaattctgattataattatgggtggga 3559
|| | || | || || | || || | |||| | |||
Db 4183 TAAAAAAAACCTATATACTAATTTATAAAATTAAGATATTTATATATATAAATATTTA 4242

Qy 3560 tacaatcgctttccactaaatatTTtaactatgatttataaatttatttcacatcgat 3619
|| || | |||| |||| | || || | |||| | |||
Db 4243 ATATATTATTATATATCTAATAATTTAAATAAAAAATTTTAAAAATTTAAAAATATAGATA 4302

Qy 3620 atttacttattaatacataaatttatcataattttatggaaattgagaccaagaacatta 3679
|| || | ||| |||| | |||| | |||| | |||| | |||
Db 4303 TAATTATATAAAATTTTATATCTCATATTTATTATTATTATTAATTTAAITTTATATAAATAA 4362

Qy 3680 agagaacaaattctataacaagacaatttagaaaaaatgtacttttaggtaatttta 3739
|| || | |||| |||| | |||| | |||| | |||| | |||
Db 4363 TATAATGATTTAATTAATTATTATATATTTATAAATTTATATATTATTGAATATTTATAT 4422

Qy 3740 gtactcttaaccaaacacaaaaattcaaatcaaatgaactaaataagataataacata 3799
|| || | |||| |||| | |||| | |||| | |||| | |||
Db 4423 ATAATATATATATATATAGAAAAATAAAATTTATTAAATAATTTTCATAAAATTTAAAA 4482

Qy 3800 cggaaaccttacttgttaattcttacattcccataattttattatgaaaaataactttata 3859
|| |||| | |||| |||| | |||| | |||| | |||| | |||
Db 4483 --AAATTCCTTAAATGTATTATTTAATAAAAAAATACTTTTTAAAAAAAATAATTTTAAT 4540

Qy 3860 ttactcgaactaaatgtgtgcacaaatttatcttaaataaagaaaaaacacttaatttt 3919
|| || | ||| |||| | |||| | |||| | |||| | |||
Db 4541 TTTTTTAAAAAATAGTAAATTAATAAAAAAAAAAAAAAAAAAAAAATGAAATATATTTAT 4600

AUTHORS:

DMU37541 19517 bp DNA circular INV 04-APR-2000
Drosophila melanogaster complete mitochondrial genome.
U37541
U37541.1 GI:1166529
Drosophila melanogaster.
Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 12511 to 12682)
Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and
Wolstenholme,D.R.
Drosophila mitochondrial DNA: a novel gene order
Nucleic Acids Res. 10 (21), 6619-6637 (1982)
83090428
2 (bases 5269 to 5695)
Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
flanking sequences and comparisons to mammalian mitochondrial tRNA
genes
Nucleic Acids Res. 11 (8), 2411-2425 (1983)
83220794
3 (bases 404 to 5272)
de Bruijn,M.H.
Drosophila melanogaster mitochondrial DNA, a novel organization and
genetic code
Nature 304 (5923), 234-241 (1983)
83245048
4 (bases 804 to 1778)
Satta,Y., Ishiwa,H. and Chigusa,S.I.
Analysis of nucleotide substitutions of mitochondrial DNAs in
Drosophila melanogaster and its sibling species
Mol. Biol. Evol. 4 (6), 638-650 (1987)
88174373
5 (bases 5268 to 13619)
Garesse,R.
Drosophila melanogaster mitochondrial DNA: gene organization and

Db 137238 ATATAAAATATATATAATAATGTATATATAAATGATATTATAATATGTAATTATTAAAG 137179
 Qy 4616 cgtaatatatagtttaataaaaaagatgggtgattgggaatgtgtgtgcattcctccat 4675
 Db 137178 TGTAAAGATGATAATTAATTATATATTATATAATAATAGTTTATTATAATATGTAATAT 137119
 Qy 4676 gcacataatggtgaatctcttgcatacatagaaattcctaaatggttatagtttatgttat 4735
 Db 137118 ATTATAAT---TAAATGTATGTTATTGTATGATTTAGTATTTAATTAATATGAATAAA 137063
 Qy 4736 agtgtatgttagtagtgaataatttttaaatgtgtatctaattgtaacatcacttggct 4795
 Db 137062 AATGTATATATATTATAGAAATAAAATATATATTTTTTATAAAATAATATATAAGAAA 137003
 Qy 4796 tgatttatgttatgttatgttttactttaatgatattgcattgtattgttaatttaaca 4855
 Db 137002 GAATATATTTTATTATAGTTATATATGTATTATATAGTGATTTTTTAAATATTATAAATA 136943
 Qy 4856 ttgcttgatcattatactcttctactattaattataaatggcactgtttgtttaaactt 4915
 Db 136942 ATATAGTCTTATATTATATTTTATATAGTTATTATAAATAATATGATTATTATAAATAT 136883
 Qy 4916 ttacaaggttaagacatgtataaatatgatgacaataataacaggttttagttcaatgt 4975
 Db 136882 ACTTAAATATATAATATATATTATATATAAATAAGTAATAATTATATTATAATTAGA 136823
 Qy 4976 tagctatcttagtatgtttatgatcttaattacatttaaacaaattccacttaaatt 5035
 Db 136822 TAATTTTATAAAATTAATAAAATTAATTATTATAGTATGTATAT-AATGAAATATAATAT 136764
 Qy 5036 tttataaataataacaataaattattgtaataataacattaaatgcaacaaaaatga 5095
 Db 136763 GATTAACTACTAATAAATATATATATTATATATAATATAATAATAAATAAATATATTATA 136704
 Qy 5096 aataaataaaaaaaatagcaaaatattgttataatattgtaataataatgtaccatat 5155
 Db 136703 ATATTAGATATATTTGTTATATATAATAGATTATAAGACGAAAAATAGTATATTTT 136644
 Qy 5156 tottaactgaatatgggt--ctaacctataatccctaaaaatttcagtttaaatatttt-- 5211
 Db 136643 AGTTATATATATGGGACTAGTAATAATATATTAAGCTGAGTTATAGTATATATATAGAT 136584
 Qy 5212 -tataacctaccatattattagaactctttttaaatattaaaaatttaattataccaat 5270
 Db 136583 GAAAATATAATAGTTAAATATAAGTATATATCTATAATATAATATATAGTATTATAGGTA 136524
 Qy 5271 ttaataaactattaattatcttaactaaaaatctaaaattttatttaacctattaataaa 5330
 Db 136523 TTGAATATATAATTATAAAATATATATTAATCAGAAAAATATATTTTGATTAGTAATTTT 136464
 Qy 5331 ttcttaattatcttactctaattttaaactctaattatcttaatttaatttaattcttaa 5390
 Db 136463 ATAATCATATAGTTATTATATATAATGATAGTTATNNNNNNNNNNNNNNNNNNNNNN 136404
 Qy 5391 ttactctaattgttaacctctctccaccagctagatgctggaccgcaatccggagatta 5450
 Db 136403 NNNATA 136344
 Qy 5451 cat 5453
 Db 136343 TAT 136341

```

RESULT 10
DMU11584
LOCUS      DMU11584      4601 bp      DNA              INV      23-JUL-1994
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.
ACCESSION  U11584
VERSION    U11584.1  GI:508026
KEYWORDS   mitochondrial DNA; A+T region; tandem repeats.
SOURCE     fruit fly.
            ORGANISM   Mitochondrion Drosophila melanogaster

```

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
 Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
 Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
 Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
 Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
 Pfeiffer,B., Poon,L., Sequeira,A., Sethi,E., Snir,E.,
 Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
 Rubin,G.M.

TITLE Direct Submission

JOURNAL Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT On Mar 8, 2000 this sequence version replaced gi:7025688.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgpe@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 133 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 829: contig of 829 bp in length
 * 830 909: gap of unknown length
 * 910 1649: contig of 740 bp in length
 * 1650 1729: gap of unknown length
 * 1730 1953: contig of 224 bp in length
 * 1954 2033: gap of unknown length
 * 2034 2661: contig of 628 bp in length
 * 2662 2741: gap of unknown length
 * 2742 3349: contig of 608 bp in length
 * 3350 3429: gap of unknown length
 * 3430 3968: contig of 539 bp in length
 * 3969 4048: gap of unknown length
 * 4049 4504: contig of 456 bp in length
 * 4505 4584: gap of unknown length
 * 4585 5411: contig of 827 bp in length
 * 5412 5491: gap of unknown length
 * 5492 6373: contig of 882 bp in length
 * 6374 6453: gap of unknown length
 * 6454 7449: contig of 996 bp in length
 * 7450 7529: gap of unknown length
 * 7530 8139: contig of 610 bp in length
 * 8140 8219: gap of unknown length
 * 8220 9315: contig of 1096 bp in length
 * 9316 9395: gap of unknown length
 * 9396 10239: contig of 844 bp in length
 * 10240 10319: gap of unknown length
 * 10320 10964: contig of 645 bp in length
 * 10965 11044: gap of unknown length
 * 11045 11648: contig of 604 bp in length
 * 11649 11728: gap of unknown length
 * 11729 12696: contig of 968 bp in length
 * 12697 12776: gap of unknown length
 * 12777 13976: contig of 1200 bp in length
 * 13977 14056: gap of unknown length
 * 14057 15045: contig of 989 bp in length
 * 15046 15125: gap of unknown length
 * 15126 15969: contig of 844 bp in length
 * 15970 16049: gap of unknown length
 * 16050 16859: contig of 810 bp in length
 * 16860 16939: gap of unknown length
 * 16940 17662: contig of 723 bp in length
 * 17663 17742: gap of unknown length
 * 17743 18767: contig of 1025 bp in length
 * 18768 18847: gap of unknown length
 * 18848 19809: contig of 962 bp in length
 * 19810 19889: gap of unknown length
 * 19890 21046: contig of 1157 bp in length
 * 21047 21126: gap of unknown length

* 21127 21826: contig of 700 bp in length
 * 21827 21906: gap of unknown length
 * 21907 23735: contig of 1829 bp in length
 * 23736 23815: gap of unknown length
 * 23816 25556: contig of 1741 bp in length
 * 25557 25636: gap of unknown length
 * 25637 26792: contig of 1156 bp in length
 * 26793 26872: gap of unknown length
 * 26873 28359: contig of 1487 bp in length
 * 28360 28439: gap of unknown length
 * 28440 29898: contig of 1459 bp in length
 * 29899 29978: gap of unknown length
 * 29979 31836: contig of 1858 bp in length
 * 31837 31916: gap of unknown length
 * 31917 33347: contig of 1431 bp in length
 * 33348 33427: gap of unknown length
 * 33428 34568: contig of 1141 bp in length
 * 34569 34648: gap of unknown length
 * 34649 35754: contig of 1106 bp in length
 * 35755 35834: gap of unknown length
 * 35835 37815: contig of 1981 bp in length
 * 37816 37895: gap of unknown length
 * 37896 39641: contig of 1746 bp in length
 * 39642 39721: gap of unknown length
 * 39722 41135: contig of 1414 bp in length
 * 41136 41215: gap of unknown length
 * 41216 42477: contig of 1262 bp in length
 * 42478 42557: gap of unknown length
 * 42558 44229: contig of 1672 bp in length
 * 44230 44309: gap of unknown length
 * 44310 45922: contig of 1613 bp in length
 * 45923 46002: gap of unknown length
 * 46003 47999: contig of 1997 bp in length
 * 48000 48079: gap of unknown length
 * 48080 49982: contig of 1903 bp in length
 * 49983 50062: gap of unknown length
 * 50063 51360: contig of 1298 bp in length
 * 51361 51440: gap of unknown length
 * 51441 53101: contig of 1661 bp in length
 * 53102 53181: gap of unknown length
 * 53182 54926: contig of 1745 bp in length
 * 54927 55006: gap of unknown length
 * 55007 56937: contig of 1931 bp in length
 * 56938 57017: gap of unknown length
 * 57018 57606: contig of 589 bp in length
 * 57607 57686: gap of unknown length
 * 57687 58632: contig of 946 bp in length
 * 58633 58712: gap of unknown length
 * 58713 60613: contig of 1901 bp in length
 * 60614 60693: gap of unknown length
 * 60694 62727: contig of 2034 bp in length
 * 62728 62807: gap of unknown length
 * 62808 65311: contig of 2504 bp in length
 * 65312 65391: gap of unknown length
 * 65392 66685: contig of 1294 bp in length
 * 66686 66765: gap of unknown length
 * 66766 68830: contig of 2065 bp in length
 * 68831 68910: gap of unknown length
 * 68911 71103: contig of 2193 bp in length
 * 71104 71183: gap of unknown length
 * 71184 72193: contig of 1010 bp in length
 * 72194 72273: gap of unknown length
 * 72274 74138: contig of 1865 bp in length
 * 74139 74218: gap of unknown length
 * 74219 76236: contig of 2018 bp in length
 * 76237 76316: gap of unknown length
 * 76317 77913: contig of 1597 bp in length
 * 77914 77993: gap of unknown length
 * 77994 80808: contig of 2815 bp in length
 * 80809 80888: gap of unknown length
 * 80889 82776: contig of 1888 bp in length
 * 82777 82856: gap of unknown length
 * 82857 85682: contig of 2826 bp in length

Query Match 3.5%; Score 192.6; DB 33; Length 86829;
Best Local Similarity 43.8%; Pred. No. 4.1e-14;
Matches 1064; Conservative 0; Mismatches 1354; Indels 9; Gaps 5;

Qy 1566 tgtgttttcttgaaaatttcgattaacataatcatgcattctcaattttggccaattg 1625
| | | | | | | | | | | | | | | | | |
Db 39027 TATTTTAAATAAATAATATAAATTATTAATAAAAATAAAAAAAATATATGTAATAGTT 39086
| | | | | | | | | | | | | | | | | |

Qy 1626 aacgttataaaattctctatgatatcctgatctgtttattaccattatatgtgtttatgct 1685
| | | | | | | | | | | | | | | | | |
Db 39087 ATATATATAATATTAAATTAAATAAAATTAATATAATAATAAATAAATAATATAT 39146
| | | | | | | | | | | | | | | | | |

Qy 1686 tgagttgaagtcaaacattgagattcatagctcaccccaatttattaatcatttcaggcaat 1745
| | | | | | | | | | | | | | | | | |
Db 39147 ATATTAATAAATAAATAAACAACAAATAAATTAAATTATTTTAAATTAATTAATAAATA 39206
| | | | | | | | | | | | | | | | | |

Qy 1746 ctgcagacttaggattggatggcgcttcaggagcttgattggttttctcacatcatatt 1805
| | | | | | | | | | | | | | | | | |
Db 39207 AAATATATTATTATTAAAAATAAATAAATTAAATATATATTATTATTAAAAATAAAAAATA 39266
| | | | | | | | | | | | | | | | | |

Qy 1806 tatttaaaataattattaattaaattttatggacttttggactgtctgactaattttcagaa 1865
| | | | | | | | | | | | | | | | | |
Db 39267 ATTAAATATATATTATTATTAATAAATAAATAAATAAATTAATATATATTATTATTAATAAATA 39326
| | | | | | | | | | | | | | | | | |

Qy 1866 ttttatttttggttttgggttttgtgaatttttttagataattttttaaatattctgcat 1925
| | | | | | | | | | | | | | | | | |
Db 39327 AATAAATTAATATATATTATTATTATTAATAAATAAATAAATAAATTAATATATATTATTATTATA 39386
| | | | | | | | | | | | | | | | | |

Qy 1926 aatttttctgtattttgaaaaggatgttcgaatttttttcaaattgaaacgatttaaga 1985
| | | | | | | | | | | | | | | | | |
Db 39387 AATAAATAAATAAATAAATAATATATATTATTATTATTAATAAATAAATAAATAAATAAATAAATA 39446
| | | | | | | | | | | | | | | | | |

Qy 1986 atttttactactgcaaaattcagaataagtggaatttggttttttagaaagattaataaagtt 2045
| | | | | | | | | | | | | | | | | |
Db 39447 TATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 39506
| | | | | | | | | | | | | | | | | |

Qy 2046 agtattacgatttttagtttgatttgggtgaaagtaagtatgtttttgaacataattat 2105
| | | | | | | | | | | | | | | | | |
Db 39507 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 39566
| | | | | | | | | | | | | | | | | |

Qy 2106 ttgacaataattaagttttctagggaataaacggaatatcttctcttttttgtaaaat 2165
| | | | | | | | | | | | | | | | | |
Db 39567 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 39626
| | | | | | | | | | | | | | | | | |

Qy 2166 tactaatgcaagaacaacaacgcttttggggagcaataatctagctttaagtagtcagt 2225
| | | | | | | | | | | | | | | | | |
Db 39627 AATATATTATATAATATATATTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 39686
| | | | | | | | | | | | | | | | | |

Qy 2226 g---taactctcaaaatctggctcataacttctaggctgagtttgcgtgctacagtagta 2282
| | | | | | | | | | | | | | | | | |


```

RESULT      6
AC005504
LOCUS       AC005504   104992 bp    DNA             HTG             01-APR-1999
DEFINITION  Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
            ***, 3 unordered pieces.
ACCESSION   AC005504
VERSION     AC005504.3  GI:4558584
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 104992)
AUTHORS     Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B.
            and Davis,R.W.
  TITLE      Plasmodium falciparum 3D7 chromosome 12
  JOURNAL    Unpublished
REFERENCE   2  (bases 1 to 104992)
AUTHORS     Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE      Direct Submission
  JOURNAL    Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Apr 2, 1999 this sequence version replaced gi:4337172.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces

```

Qy 2278 taqtaagtctatagaaaacttacctgacaaaacgacatgacgtcagggtcgaatctacaac 2337

U49822 *Saccharomyc*
Z92859 *Caenorhabdi*
X95275 *P.falciparu*
AC007708 *Homo sapi*
AC008132 *Homo sapi*
AC004470 *Homo sapi*
AC004617 *Homo sapi*
U49822 *Saccharomyc*
AL013745 *Plasmodium*
AE001369 *Plasmodium*
AL034560 *Plasmodium*
AC006280 *Plasmodium*
AC008132 *Homo sapi*
AE001368 *Plasmodium*
Continuation (2 of
X95276 *P.falciparu*
AL034558 *Plasmodium*
AC004617 *Homo sapi*
AL035477 *Plasmodium*
AE003846 *Drosophil*
AC004153 *Plasmodium*

```

RESULT 1
GBU34401
LOCUS      GBU34401      1699 bp      DNA                      PLN      01-JAN-1996
DEFINITION Gossypium barbadense FbLate-2 gene, complete cds.
ACCESSION  U34401
VERSION     U34401.1  GI:1143223
KEYWORDS
SOURCE      sea-island cotton.
ORGANISM    Gossypium barbadense
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
             eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1 (bases 1 to 1699)
AUTHORS     Rinehart,J., Petersen,M. and John,M.E.
TITLE       Tissue-specific and Developmental Regulation of Cotton mRNA,
             FbLate-2: Promoter Studies in Transgenic Plants
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1699)
AUTHORS     John,M.E.
TITLE       Direct Submission
JOURNAL     Submitted (21-AUG-1995) Maliyakal E. John, Fiber Technology,
             Agracetus, 8520 University Green, Middleton, WI 53562, USA
FEATURES
  source
    1..1699
    /organism="Gossypium barbadense"
    /strain="Sea Island"
    /db_xref="taxon:3634"
    /clone="FbL2-82A"
  mRNA
    369..1585
    /gene="FbLate-2"
  gene
    369..1585
    /gene="FbLate-2"
  CDS
    379..1380
    /gene="FbLate-2"
    /codon_start=1
    /protein_id="AAA84881.1"
    /db_xref="GI:1143224"
    /translation="MIGSHTVSTAARRLFETQTTSSSLPQLASKYEQEESYEYKEP
    KQPKYDEEYPKHEKPEIHKEEKQKPKQHEEYHESHKSKEHEEYQKEKPEFPKLEKPI
    EHKHEVEYPKILEYKENQDEGKEHKHEEYHESRESKEHEEYKEKPEFPKLEKPEI
    EKHEVEYPEIPEYKEKQDEGKEHKHEECHKSHEESKEHEEYKEKPNFPKGEKPKHEH
    HEVEYPKIPEYKEKQDEGKEHKHDECHESHELAKHEEYKEKPNFPKGEKPKHEKH
    VEYPKIPEYKEKQDEGKEHKHEFQKHEKEEKKAPKPKAEYSEWPKSMFTQSGSGTKP
    polyA_signal
    1448..1454
    /gene="FbLate-2"
BASE COUNT      661 a      328 c      328 g      382 t
ORIGIN

```

Qy 3858 tattactcgaactaaatgtgtgcacaaattattattctaaataagaa--aaacacttaat 3915
|||||
Db 1 TATTACCTGAGCCAAATGCTCTCACAAACTATTATCCAAAAAAAATGTTGAATATAAT 60
|||||
Qy 3916 ttttataacattttttcatatatttgaagattatattttgtatattttagtataaaatat 3975
|||||
Db 61 TTTTATAACATTTTTTCATATATTGCAAGATTATATTTTGTATATTACGTAAAAATAT 120
|||||
Qy 3976 ttgacatagattgagcacctttcttaacaataatcccaccataagtcaagtatgtagatgag 4035
|||||
Db 121 TTGACATAGATTGAACACCTTCTTAACATAATCCCACCATAAGTCAAGTATGTAGATGAG 180
|||||
Qy 4036 aaattggtacaacaacagtggggccaaatcccaccaaaccatctctcattctctcttata 4095
|||||
Db 181 AAATTGGTACAAACAGTGGGGCCAAATCCCACCAAAACCATCTCTCATCTCTCTCTATA 240
|||||
Qy 4096 aaaggcttgctacacatagacaacaatccacacacaaatacac----gttctttttttt 4150
|||||
Db 241 AAAGGCTAGTTACACATACACAACAATCCACACAAAATACACTCAAAATCTTTTGCTTT 300
|||||
Qy 4151 ctattt-gattaaacctggctcatagcattcgtacccttttcttcttttccaactttta 4209
|||||
Db 301 GTATTTCGGTTAACCATGGCTCATAACACTCGTCCACCTTTCTTCCTTTTCCAACTTTTA 360
|||||
Qy 4210 ctcataagtgtctcactagtacggttagccacactgttttcggcagcggttcgacgttta 4269
|||||
Db 361 CTCATTAGTGTCTCACTAATGATCGGTAGCCACACCGTCTCGACAGCGGCTCGACGTTTA 420
|||||
Qy 4270 ttcgagacacaaagcaacctcatcagagctcccacaatttggttccaaatacgaagacag 4329
|||||
Db 421 TTCGAGACACAAACACCTCATCGGAGTTGCCACAATTAGCTTCAAAATACGAAAAGCAG 480
|||||
Qy 4330 --agagtctgaatcgaaaaagccgaataatacaaacagcgaagatcatcgaagagtactca 4387
|||||
Db 481 GAAGAGTCTGAATATGAAAAGCCGGAATACAACAGCCAAAGTATGACGAAGAGTACCCA 540
|||||
Qy 4388 aaacttgagagacctgaatgcaaaaggaggaaaaacaaaaccttgcaaacagcatgaa 4447
|||||
Db 541 AAACATGAGAAGCCTGAAATTACAAGGAGGAAAAACAAAACCGTGCAAGCAACATGAA 600
|||||
Qy 4448 gagtaccacgagtcacacgaatcaaaaggagcaaaaagagtacgagaagaanaattctga 4506
|||||
Db 601 GAGTACCACGAGTTCACACAATTCGAAGAGGACGACGAAGAGTACGAGAAAGAAAAACCCGA 659
|||||

Query Match 4.8%; Score 265.4; DB 5; Length 1283;

COMMENT - Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR08K08"
/note="end : TET3"
BASE COUNT 395 a 120 c 103 g 334 t 149 others
ORIGIN

Query Match 2.1%; Score 113.2; DB 122; Length 1101;
Best Local Similarity 42.9%; Pred. No. 5.9e-08;
Matches 268; Conservative 74; Mismatches 272; Indels 10; Gaps 3;

Qy 3323 aataaactttaaattacaagcataatattaaatttgaatcaattttttttctat 3382
|| || :||| | :| | | | | :| :|| | |
Db 488 AAAAAATWTTAAAAAAWAAAAACCTTTAATAAATAAAAAAATAAAAAAATTT 547
Qy 3383 tattttaattaatttagtctatttttcaaaataaaatttaatttaataaaataatt 3442
| : : : || | | | | | | | | | | | | : ||
Db 548 TTTWAMTTTATTAMATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 607
Qy 3443 ttcccttaattgtgaacaactcatgttataacttcaaaattataagttattattac 3502
|| | | | | | | | | | | | | : | | | | : || |
Db 608 TTTTTTTTTTTTTTTTTTTTTTTATTTTWTATTTTWTATTTTAAATAAATTTTAT 667
Qy 3503 tgaatgatttatttagtattataattctgattataattatggtgggatacatgctt 3562
| : | : || | | | | | | | | | | | | : | :
Db 668 TTANNAAWATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 727
Qy 3563 tccactaaatttttaactatgatttataaatttttcaacatgattattacttatt 3622
| : | : | : || | | | | | | | | | | | | : | : |
Db 728 TATA-TAWTWTATAAATAATATAAAAAAATAAATTTTATTATATAAATAATATAT 786
Qy 3623 aatacataattttatcataatttttgaaattgagaccaagaacattagaagaacaa 3682
: || | | | : || : || | : | | | | | : : | | | : |
Db 787 WATTATATATNWNWATAWTTTWTAAATTTWATAWTTAAATWTTAAATTTTAW 846
Qy 3683 tctataacaagaacatttagaataaattgactttttaggttaatttaagttactctta 3742
|| | | : || | | | | | | | | | | : | : || | | : | :
Db 847 AATAAAAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 904
Qy 3743 caaacacaaaaattcaaatcaaatgaactaaataagataataacatacgcgaacatt 3802
: | | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 905 AAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 957
Qy 3803 acttgtaattcttaccataattttattatgaataaataactttattactcgaact 3862
| : : || | | | | | | | | | | | | | | | : |
Db 958 AAWAATAATTTTATATWATWATTAATAAATAAATAAATAAATAAATAAATAA 1017
Qy 3863 aaatgtgtgcacaaattattatctaataaagaacaaacttaattttataacatttt 3922
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1018 WWAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1077
Qy 3923 tcatattttgaagattatttt 3946
: : : || | : : | | |

Db 1078 WAWCWTATCCNCWCYATTTTTT 1101

RESULT 13
CNS009G1/c
LOCUS CNS009G1 876 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19J14 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053529
VERSION AL053529.1 GI:4935018
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 876)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source Location/Qualifiers
1. .876
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19J14"
/note="end : TET3"
BASE COUNT 335 a 54 c 57 g 325 t 105 others
ORIGIN

Query Match 2.0%; Score 112.6; DB 122; Length 876;
Best Local Similarity 42.4%; Pred. No. 7.4e-08;
Matches 229; Conservative 54; Mismatches 255; Indels 2; Gaps 1;

Qy 4901 tttaaactttttacaagttgaacatgtataaatatgacataataattacaagtttta 4960
| | | | : | | | | : |||| | : || | | |
Db 875 TATATAATANAATAATNTAATAAAAAATATWATATATAATATAWATWATATATTA 816
Qy 4961 gttcaatgttagctatcttagtattgattgatcttaattacatttaacaattcc 5020
| | | | | | | | | | | | : | | | : : | : | :
Db 815 --TATANTATATATATATTAATWATAATTAATANAATAAATAAATAAATAAATAA 758
Qy 5021 acttaaaattttaataataataacaaataattattgtaataataacatttaaatgcaac 5080
| : | : |||| | | : | : | | | | | | | | | |
Db 757 NTATWATNNTAATAATATATATATATATATATTTTNTATAAAAAATATATATATA 698
Qy 5081 aaaaaatgaataataataataataagcaataattgtataatattgtaataataat 5140
: | | | : || | | | | : | | | : || | | | | : | : | : | : |
Db 697 WNAATATTTAAWATATWAAAAATATATWNTAAWATATATATATATWNTAAATAA 638
Qy 5141 gtaccattcttaactgaataggtctaacctataatccctaaatttcagtttaaat 5200
| | | | | : | | | | | : | | | | | : | : | | | : | : | : | : |
Db 637 TTTTATATTTTTTATTTTATATATAAATTTTATATAAATTAATTAATTAATTTTAT 578

RESULT 10
CNS0021J

```
Qy      3119 acagaataatgtttaaataagaaattaaaataaggctggcctggtcacacacacacaaaaaaa 3178
        | | | | |   |||| | | | | | | | | | | | | | | | | | | | | | |
Db       270 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA CAAAAAAAAAAAA 329

Qy     3179 ctaatgttggttggtgaattttatattacggaatgtaattattatttttaaat 3238
        | | | | |   | | | | | | | | | | | | | | | | | | | | | |
Db      330 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA CAAAAAAAAAAAA 389

Qy     3239 tatgtttatttagattcttaatttttggagcattccatactataatttcgtaacataa 3298
        | | | | |   | | | | | | | | | | | | | | | | | | | | | |
Db      390 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA CAAAAAAAAAAAA 449

Qy     3299 ttaasatatagtaataaagtgtaattaactttaattacaagcataatattaaat 3358
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      450 AAAAAAAAAAAAAAAAAAANAANAAAAAAAAAAAAAAAAAAAAAAAAA ANAAN 509

Qy     3359 gaatcaaatattttttattctatttttaatttaatttagtctatttttcaaaataa 3418
        | | | | | | | | : | | | : | | | | | | | | | | | | | | |
Db      510 AAAAAATATAATTWTWTTTTTTTAAATWTTTTTTTTTTTTTTTTTTTWTTAATTTAA 569

Qy     3419 atttaaatctaataaaaataatttttcttaagtgaacaactcatgttatacttca 3478
        | | | | : | | | | | | : : | | | | | | : | | | | | | |
Db      570 TTTTAAAWAAATTTAATAAAAWTWTTTWTWTTTTTAAATWA WAAAAAAAWTTTAA 629

Qy     3479 aaattataagtattattatttaccttgatgattttattattagtataaattctgattat 3538
```


Qy 3601 caacatcgtatatattacttattataacataaatttatcataattttatggaattgagacc 3660
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 807 AAATATTAAATTTTTTAAAAATAATTTTATTNTTATTTNTTTATATTTTTTATATT 866

Ov 3751 aaaattcaaatcaaatgaactaaataagataatataacatacgaacatcttacttqtaa 3810

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query				Description
	Score	Match	Length	ID	
c	1	145.6	2.6	1101 122	CNS00EVL AL069706 Drosophil
	2	142	2.6	1101 122	CNS00EVL AL069706 Drosophil
	3	126.2	2.3	1101 122	CNS0039G AL063921 Drosophil
c	4	121.6	2.2	1201 123	CNS0167M AL106396 Drosophil
	5	119.8	2.2	1101 122	CNS00E07 AL069440 Drosophil
c	6	118	2.1	1101 122	CNS00E07 AL069440 Drosophil
	7	117	2.1	1187 120	B11102 B11102 F19C22-T7 I
c	8	116.8	2.1	1101 122	CNS00KAE AL077628 Drosophil
c	9	115.2	2.1	1187 120	B11102 B11102 F19C22-T7 I
	10	115	2.1	1101 122	CNS0021J AL061936 Drosophil
c	11	114.2	2.1	1101 122	CNS003BD AL064091 Drosophil
	12	113.2	2.1	1101 122	CNS003BD AL064091 Drosophil
c	13	112.6	2.0	876 122	CNS009G1 AL053529 Drosophil
c	14	111.2	2.0	1101 122	CNS0021J AL061936 Drosophil
	15	110.8	2.0	1201 123	CNS0167M AL106396 Drosophil
c	16	110.2	2.0	1225 123	CNS0161D AL106171 Drosophil
	17	109.8	2.0	1101 122	CNS0039G AL063921 Drosophil
c	18	109.8	2.0	1101 123	CNS0145U AL103740 Drosophil
	19	109.6	2.0	1101 122	CNS00EPO AL069493 Drosophil
c	20	109.4	2.0	1101 122	CNS00BO1 AL057419 Drosophil
	21	108.2	2.0	1101 122	CNS000B8 AL063632 Drosophil
c	22	108.2	2.0	1101 122	CNS003BB AL064089 Drosophil
	23	108	2.0	734 122	CNS010MP AL099163 Drosophil
c	24	108	2.0	836 122	CNS01100 AL099642 Drosophil
	25	107	1.9	935 120	B10881 B10881 F24H6-Sp6.1
c	26	107	1.9	1101 122	CNS00EPO AL069493 Drosophil
	27	106.8	1.9	718 102	AQ416310 RPCI-11-1
c	28	106.2	1.9	1101 122	CNS001FB AL060732 Drosophil
	29	106	1.9	828 113	AQ739398 HS_5482_B
c	30	105.8	1.9	836 122	CNS01100 AL099642 Drosophil
c	31	105.2	1.9	1101 122	CNS004ZW AL055440 Drosophil
	32	105	1.9	928 122	CNS00DKY AL071865 Drosophil
c	33	105	1.9	1101 122	CNS00KAE AL077628 Drosophil
	34	104.4	1.9	1101 122	CNS003BB AL064089 Drosophil
c	35	104.4	1.9	1101 122	CNS003DX AL064587 Drosophil
	36	103.8	1.9	1101 122	CNS000B8 AL063632 Drosophil
c	37	103.6	1.9	1101 123	CNS0145U AL103740 Drosophil
c	38	103.2	1.9	1101 122	CNS003DQ AL064580 Drosophil
	39	102.4	1.9	890 93	AQ026918 AQ026918 CIT-HSP-2
c	40	101.8	1.8	876 122	CNS009G1 AL053529 Drosophil
	41	101.6	1.8	1101 122	CNS00PYG AL071206 Drosophil
c	42	101.4	1.8	928 122	CNS00DKY AL071865 Drosophil
	43	101.4	1.8	996 122	CNS00FUH AL071063 Drosophil
c	44	101.2	1.8	905 122	CNS00KHX AL077798 Drosophil
	45	101	1.8	828 113	AQ739398 HS_5482_B

ALIGNMENTS

```

RESULT 1
CNS00EVL
LOCUS      CNS00EVL      1101 bp      DNA      GSS      04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit

```

fly), genomic survey sequence.

ACCESSION AL069706

VERSION AL069706.1 GI:4949849

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM *Drosophila melanogaster*
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; *Drosophila*.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley *Drosophila* Genome Project (BDGP).
The BDGP is constructing a physical map of the *Drosophila*
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP *Drosophila*
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of *Drosophila* DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29B23"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match 2.6%; Score 145.6; DB 122; Length 1101;
Best Local Similarity 38.7%; Pred. No. 9.2e-13;
Matches 249; Conservative 130; Mismatches 260; Indels 5; Gaps 2;

[illegible]

Db 3812 TTTGTAAGTTTAAATAATCATCATTTTGTATTCTTTTTATCGATAATATT--GG 3756

Qy 3392 taatttagtctattttttcaaaaataaaatttaaactcaaaataaaataattttctctaa 3451
| | | | | | | | | | | | | | | | | | | | | |

Db 3755 TGGATTGTCATTTTTTTAGGAATTCATTTTATATGTATTATCACTTTTTTGTATTAT 3696

Qy 3452 tgttgaacaactcatgttatacttcaaaattataagtattatattacccttgatgatt 3511
| | | | | | | | | | | | | | | | | | | | | |

Db 3695 TCATAATTATTTTGAAAATAGTAATACCGTGTAAATATACAAACCTAAAAATGTTATTA 3636

Qy 3512 atttattagtatataaattctgattataattatggtgggatacaatcgctttccactaaa 3571
| | | | | | | | | | | | | | | | | | | | | |

Db 3635 ACITTTAAGTTTTTTTTTTTTTTTTTTTTTTTATATTAGAATAATTGTAACCATTTAAA 3576

Qy 3572 tattttaactatgatttataaatttatttcaacatcgatatattacttattaatacataa 3631
| | | | | | | | | | | | | | | | | | | | | |

Db 3575 TATTGGAGTATAAATAATATATATATTATAAC--GAGACAATTAGTAAAAAAAATAG 3518

Qy 3632 tttatcataatttttggaattgagaccaagaacattagaagacaattcttataaca 3691
| | | | | | | | | | | | | | | | | | | | | |

Db 3517 TTAAAAAAAATCGTTAAAAAAAATATGAAAATAAATGGATATATAATTGAATGAATAA 3458

Qy 3692 aagacaatttagaaaaaaatgtacttttaggtaatttttaagtactcttaaccaaacacaa 3751
| | | | | | | | | | | | | | | | | | | | | |

Db 3457 CATAAAA--AGATGACAATTTATCAAACTGTTATTTAAAAATAACTTAATCATACAAAA 3401

Qy 3752 aaattcaaatacaatgaactaaataagataatataacatacggaaacatcttactgtaat 3811
| | | | | | | | | | | | | | | | | | | | | |

Db 3400 AAAAAGGAACAAAAACAGGAAAAAGGAATAAGCTGTAAGAAACACAAACAATTTAAAGA 3341

Qy 3812 cttacattcccataattttattatgaaaaataatcttattactcgaaactaaatgttgt 3871
| | | | | | | | | | | | | | | | | | | | | |

Db 3340 CAAGCGAATTTATGAATTTATTATTAAAGGTATATTATATATATATCATAGATAATAT 3281

Qy 3872 cacaaatttattatctaaataaagaaaaaacacttaatttttataacattttttcatatatt 3931
| | | | | | | | | | | | | | | | | | | | | |

Db 3280 TTAAAAGCAAAAAAAGACAAACATATATCAGATTTGTATGTAATATAAGAAATAATAAG 3221

Qy 3932 tgaagagattatattttgtatatttcagtaaaaaattatt 3968
| | | | | | | | | | | | | | | | | |

Db 3220 TGGATATATACGTTTATTAAAAATAGTTCAGTCATT 3184

RESULT 15

PCT-US92-00018-1/c

; Sequence 1, Application PC/TUS9200018

GENERAL INFORMATION:

APPLICANT: Hoffman, Stephen L.

APPLICANT: Charoenvit, Yupin

APPLICANT: Hedstrom, Richard

APPLICANT: Khusmith, Srisin

; APPLICANT: Rogers IV, William O.

TITLE OF INVENTION: Protective malaria sporozoite surface protein

; TITLE OF INVENTION: immunogen and gene encoding

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

ADDRESSEE: A. David Spevack

STREET: NMRDC Building 1 T-12 National Naval

STREET: Medical Center

CITY: Bethesda

; STATE: MD

; COUNTRY: USA

ZIP: 20814-5044

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```
; SOFTWARE: PatentIn Rel
```

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: P

; FILING DATE: 199201

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

```

; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; LIBRARY: Py-lambdagt11-2-7 kb genomic expression
; CLONE: Py10.1111
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 718..3195
; OTHER INFORMATION:
PCT-US92-00018-1

```

Query Match 1.4%; Score 79; DB 6; Length 4673;
Best Local Similarity 46.8%; Pred. No. 1.2e-05;
Matches 354; Conservative 0; Mismatches 395; Indels 8; Gaps 3;

Qy 3212 atgtaatatattatattttaaaataaaatattgttatttagattcttaatatattttggagcat 3271
| | | | | | | | | | | | | | | | | | | | | |
Db 3932 AAGGATGATTAATACTTCAAAAAATCATGAGCCAACCTTTAGATATTCTCTTTTTAACATT 3873

Qy 3272 tccatactataatttcgtaacataatattaaaatatagtaataaagtgtaattaactt 3331
| | | | | | | | | | | | | | | | | | | | | |
Db 3872 CCATCATTTTTTTTTATCACACTTTTTAGTTCATAAAACTTAAGACCATTATTTTATGT 3813

Qy 3332 taaattacaagcataatattaaattttgaatcaattaaattttattctattattttaat 3391
| | | | | | | | | | | | | | | | | | | | | |
Db 3812 TTTGTAAATTTTAAATAATCACATTTTGGTTATTCTTTTTATCGATAATATT---GG 3756

Qy 3392 taatttagtctattttttcaaaataaaattaaatctaaaataaaataatttttctctaa 3451
| | | | | | | | | | | | | | | | | | | | | |
Db 3755 TGGATTGTCTATTTTTTTAGGAATTCATTTTATATGATTATACACTTTTTTGTTTTAT 3696

Qy 3452 tgttgaacaactcatgttatacttcaaaattataagttattattacccttgatgatt 3511
| | | | | | | | | | | | | | | | | | | | | |
Db 3695 TCATAATTATTTTGAATAATAGTAATACCGTGTAAATATACAAACCTAAAAATGTTATTA 3636

Qy 3512 atttattagttatattattctgattataattatgggtgggatacaatcgctttccactaaa 3571
| | | | | | | | | | | | | | | | | | | | | |
Db 3635 ACTTTTAAGTTTTTTTTTTTTTTTTTTTTTTTATATTAAGAATAAATGTGAACCATTAATA 3576

Qy 3572 tattttaactatgattttataaatttatttcaacatcgatatattacttattaatacataa 3631
| | | | | | | | | | | | | | | | | | | | | |
Db 3575 TATTGGAGTATAATAATATATATATTATAAC--GAGACAATTAGTTAAAAAAAATAG 3518

Qy 3632 tttatcataattttatggaattgagaccaagaacattaagagaacaattctataaca 3691
| | | | | | | | | | | | | | | | | | | | | |
Db 3517 TTAATAAAAAAATCGTTAAAAAAAATATGAAAATAAATGGATATATAATTGAATGAATAA 3458

Qy 3692 aagacaatttagaataaaattgtacttttaggtaatttttaagtactcttaaccaaacacaa 3751
| | | | | | | | | | | | | | | | | | | | | |
Db 3457 CATAAAA---AGATGACAATTATCAAACTGTTAATTAAAAATAACTTAATCATACAAAA 3401

Qy 3752 aaattcaaatcaaatgaactaaataagataataataacatcggaacatcttaactgttaat 3811
| | | | | | | | | | | | | | | | | | | | | |
Db 3400 AAAAAGGACAAAAACAGGAAAAAGGAATAAAGTGTAAGAGAAACACAACAACTTTTAAGA 3341

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Qy 4782 acctggcttgatttatgttatgttatgtattttactttaatgatattgcagtgtattga 4841
| ||| || | | | | | | | | | | | | | | | | | | | |
Db 14 AATTGTGTTTTTATAAATTAATATTTTATAAATTAATATTTTATAAATTAATATTTTATA 73

Qy 4842 atttaacattgcttgcattataactcttctactattaattataaatggcactgtttgt 4901
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 74 ATTAAATATTTTATAAATTAATATTTTATAAATTAATATTTTATAAATTAATATTTTATA 133

Qy 4902 ttaaacctttttacaagttaagacatgtataaataatgatataattacaagttttag 4961
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 134 ATTAAATATTTTATAAATTAATATTTTATAAATTAATATTTTATAAATTAATATTTTATA 193

Qy 4962 ttcaatgtagctatcttagtatgttatgtatgaccttaattacatttaacaaattcca 5021
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 194 ATTAAATATTTTATAAATTAATATTTTATAAATTAATATTTTATAAATTAATATTTTATA 253

Qy 5022 cttaaaattttaataataa-----taacaaataattattgtaataataacattaat 5075
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 254 ATTAAATATTTTATAAATTAATATTTTATAAATTAATATTTTATAAATTAATATTTTATA 313

Qy 5076 gcaacaaaaaatgaaataaaaaataaaatagcaataaattggtataatattgtaatat 5135
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 314 ATTACATATTTTATAAATTAATATTTTATAAATTAATATTTTATAAATTAATATTTTATA 373

Qy 5136 aatatgtaccatattcttaactgaatatgggtctaacctataatccctaaaatttcagtt 5195
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 374 ATTACATATTTTATAAATTAATATTTTATAAATTAATATTTTATAAATTAATATTTTATA 433

Qy 5196 taaatattttttacctgccatattattagaactcttt----ttaaatatattaaaaatt 5251
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 434 ATTACATATTTTATAAATTAATATTTTATAAAGTATTATAAATTAATATTTTATAAATTC 493

Qy 5252 taattataccaatttaatttaaactattaattatcttaactaaaatctaaattttattt 5311
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 494 AAGTATTATAAATTAATATTTTATAAATTAAGTATTATAAATTAATATTTTATAAATTC 553

Qy 5312 aacctattaattaaattcctaattatcttatcttaattaaaa 5353
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 554 AATATTTTATAAATAGTTAAAAACGACGAGGAAAAATTTAAA 595

```

RESULT 13
US-08-883-795A-36/c
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; . APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada

```

```

RESULT 10
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19

```

Query Match 1.5%; Score 85; DB 1; Length 5852;
Best Local Similarity 46.1%; Pred. No. 1.2e-06;
Matches 403; Conservative 0; Mismatches 460; Indels 12; Gaps 3;

```

Qy 3112 actgtcacagataatgtttaaatgaataaaataaggtggcctgtgcacacacaaa 3171
    | | | | | | | | | | | | | | | | | | | | | |
Db 1381 AGTAGCACATAAATATGTATAAACCAAAAAAATTGATTAAAGATAAGTTATATGTTGT 1440

Qy 3172 aaaaaaactaattgttggtggtgaattttatattaccggaatgtaattattatattttaaa 3231
    | | | | | | | | | | | | | | | | | | | | | |
Db 1441 ATTTAATAAAATAGTTTAGTTTAAAAATTTTATATCATTTTTTAAAAAATGAAAATGTTTG 1500

Qy 3232 ataaaaattatgttatttagattctttaatattttggagcattccataactataatttcgtaa 3291
    | | | | | | | | | | | | | | | | | | | | | |
Db 1501 AAAAAAAAAAATTTTTTTTTTTTTTTTCAACGGGACGATGTAATATCATATGATTCAAAT 1560

Qy 3292 cataattataaaatatagtaataataagtgtaattaactttaaattacaagcataatatt 3351
    | | | | | | | | | | | | | | | | | | | | | |
Db 1561 TAAAGTTATTAAACAATATGTAAAAATTATAAAAAACAACTAGTTATAATTACTTTTC 1620

Qy 3352 aaattttgaatcaattaattttt---atttctatttttaattaatttagtctattttt 3408
    | | | | | | | | | | | | | | | | | | | | | |
Db 1621 CCCCTCTTTTTTTTTTTTTTTTTTGTGTCATGACACTTTTTTTTTTGTGTCATGACACTTTT 1680

Qy 3409 tcaaaaataaaatttaaatctaataaaaaataattttcccttaagtgtgaacaaactcatg 3468
    | | | | | | | | | | | | | | | | | | | | | |
Db 1681 TTTTAAAAAAAAAAAAAAAAAATGTTAAATACTATTTGATGACATTCATTTTTCCTAGTT 1740

Qy 3469 ttatacttcaaaattataagtattattattacottgatgatttatttattagt--atatt 3526
    | | | | | | | | | | | | | | | | | | | | | |

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 1.8%; Score 100.4; DB 4; Length 19124;
Best Local Similarity 44.1%; Pred. No. 3.1e-09;
Matches 877; Conservative 0; Mismatches 1091; Indels 22; Gaps 10;

Qy 1962 attgaacgttttaagaatttttactactgcaaatcagaataagtgattgttttttag 2021
Db 4660 ATTCAATATTAGAGATTATGTAATATTGTTATGTATGTAATATATATTAATATAATT 4719
Qy 2022 aaagattaaataagtttagtattacgatttttagtttgattggtggaagtaagtgtatg 2081
Db 4720 GTTTTTTAGTATGTATGTTATCTAATAATATATCATATGTAGTCATAGTCAATGA 4779
Qy 2082 ttttgacataattatttgacaataaataagttttctagggaataaacggaaatattcttc 2141
Db 4780 ATATAAAATATGCTATATTATATATTGTATATATTAATAAGTAACACAGA-ACATTA 4838
Qy 2142 tctctttttgtaaaattactaatgcaagaacaaacacggttttgggggcaataatctta 2201
Db 4839 TATATAGTAATAAATAGAGAAATATATATTTTTATGTTATATATTATTAGTTATTATA 4898
Qy 2202 gctttaagtagtcagtgtaactctcaaaatctgggtcataactcttaggtgagttgtgtg 2261
Db 4899 AAGGGGAAATTCATAATATTTTATGAAATTTTGTATATGATATAGTTAAGTTAAAA 4958
Qy 2262 tgctacagtagtaagtctatagaaactactgacaaaacacacacgagtcagggtcgaa 2321
Db 4959 AAAAAAAAAAACAAGAAAAATGGAAGCATAAAAATGTTACTGTAATAGGATAAAA 5018
Qy 2322 tctacaacttttctttttcttcaattacatagtggtgattcaagttccgatctataat 2381
Db 5019 TATATTATATAAAATGTTTATTATCTTAAAAAGGTTCTATTATAACATTAATAAAAA 5078
Qy 2382 aatttattacgattttatcaatttcaattacatttatcatctattataaataaagtc 2441
Db 5079 TTGTGCCATTTTATAAATAAATACTACATTTACATAATGAAATTCGATTTTGTGTTT 5138
Qy 2442 gttcaattcagtttttcgaaagttcccaaaattttgaattttttaaattttccctaa 2501
Db 5139 TTTTGATGAATATTATGGACTAATTATTATATGTGAATGCGTTCTATATAATAATA 5198
Qy 2502 aaccgaaatagttatattctttcaatttaagttttcttttcaatccgattt-caatttc 2560
Db 5199 ATTTTATTTAAAAAATGAAAAATAAGAAATAAATATCTGATTTTGTAGTTCCAAATAGC 5258
Qy 2561 atccttttataactctctattatctataattacataaatttcaatttaatttgaatata 2620
Db 5259 TTAATATAATTATGGACTCATATATATATATATCTTACACAAAGTAATAAGTAA 5318

Qy 2621 ttacacttttagtcctcaagttcaaaactataaattttcacttttagaattaatcattttt 2680
Db 5319 ATATTATTTTAACTTTAATAAGGAAAAATAAATAAATAAAGAA---TACTGAATAA 5375
Qy 2681 cacatctaagcatcaaatttaaccaaatgacacaaatttcattgattagtagtaagct 2740
Db 5376 TAAGTCATATTATACATTTTTTAAAAATGTAACATAATTACAAATACGTAACTGATTA 5435
Qy 2741 tttgagttctcaaaacataaaaaattacaaaaaaacaaacttaaatcatttatcaat 2800
Db 5436 TAGAAATAATAAGAATTTAATATTAAGGATAAATAAATAATTAAAAATATATTTTTT 5495
Qy 2801 ttgaacacaaagcttgccggaatgtagagcttaaaatgcttctttgtttctttt 2860
Db 5496 ATGCAATTATGTTATATTATATTATATTACATGATTA-GTTTTGAAAAATATTA 5554
Qy 2861 tgttgcaaacggttgagagagaggaagaatgaagattgaccattttttattatgttt 2920
Db 5555 AATATCATATAATAATAAATAATAGTTAAAAATAAGTATTTTATACAAATACTAACT 5614
Qy 2921 taacataataatttaataatttaataataatacttttggtgaatgtagcagtgaggag 2980
Db 5615 TATAAGTATATCATATAATATATATATATATATTTATGTGTTTTGATTGGGTGTT 5674
Qy 2981 atacgtaaagtattttaacattatactttttgcaagcagttggtggtctaccgaaggt 3040
Db 5675 ATAAGGCTATAAGTATATATGTTGTTGTTTATATATTTATATGTAATAGATACATAT 5734
Qy 3041 gatcaaa-----gtttgagctgccttcaatgagccaaattttgcccataatggataaa 3093
Db 5735 AAGTTAATATATTTTATTTGTGTATATGCTGTGTTAAGATAGATATGCATTACAGTTAAG 5794
Qy 3094 ggcaatttggttagttcaactgctcagagaataatgtaaaatgaatttaaaataaggtg 3153
Db 5795 GGTATAGTTTTTTTTTTTTTTTTTTGTACATATATATAAATAAGATAACTAACATA 5854
Qy 3154 gcctggtcac 3213
Db 5855 TGCATATTACAAGAATAATATTGTATAAATAATATATATATATATATATAAAGACAT 5914
Qy 3214 gtaattattatttttaaaataaaattatgttatttagattcttaatttttgagcattc 3273
Db 5915 -TAAACTATACTAATAGGTAATTAGTTTATATATCATCTTTTATTATTATAATTTT 5973
Qy 3274 catactataatttcgttaacataataataaataatagtaataaagtgtaattaacttta 3333
Db 5974 TTTTGTCTTACTCTTGTCTCTCTTTTGTATTATAAATAAACAATAAACAATA 6033
Qy 3334 aattacaagcataatatttaattttgaatcaattaattttttattctatttttaatta 3393
Db 6034 TCAGTATTGGAATATAAATAAATTTTCTACATATATGCATATATATATATATATA 6093
Qy 3394 atttagctattttttcaaaataaaatttaatactaaataaaataattttt---cctta 3450
Db 6094 TAT 6153
Qy 3451 atgttgaacaaactcatgttatacttcaaaattataagattattatttacccttgatgatt 3510
Db 6154 ATTTTATATATTTTAGTATATACTTTAAAGATATTATTAATATTATATAGTAGCATAT 6213
Qy 3511 tatttattagttatattaattctgattataaattatggtgggatacaatcgctttccactaa 3570
Db 6214 ATGATTTTATATATAAATAAATTTTCATTATATAAATAATAGAACATGAACATTTT 6273
Qy 3571 atatttttaactatgatt-tataaattttttcaacatcgatattttactattataatcat 3629
Db 6274 ATTAATAACTCATATTTGAATATATATATTATAATGTGATTTTACTATTTTTTTAT 6333
Qy 3630 aatttatcataattttatggaattgagaccaagaacattagagaacaaattctataa 3689
Db 6334 ATTATACAATAAATTTTGAATTCATAAATGCATGAATACATAAATAAATAACAACAA 6393

Qy 4244 gcagcggctcgagctttattcgagacacaaagcaacctcatcagagctcccaaatggct 4303
Db 133 TCAGCGGCTCGACATTATTTCAGACACAAACACCTCATCAGAGCTGCCACAATGGCT 192

Qy 4304 tcaaaatacgaagcagcaagagctgtgaatacgaagccagaatacaaacagccaaag 4363
Db 193 TCAAAATACGAAAAGCACAAGAGTCT-----GAATACAAACAACCAAAA 237

Qy 4364 taccagcaagagtactcaaaacttgagaagcctgaaatgcaaaaggaggaacacacaa 4423
Db 238 TATCAGCAAAAGTACCCAAAACATGAGAAGCCTAAATGCACAAGGAGGAAAAACAAAA 297

Qy 4424 cccgtcaaacagcatgaagagtaccagagtcacacgaatcaaggagcaaaagagtag 4483
Db 298 CCCTGCAACATCATGAAGAGTACCAGAGTCACGCGAATCGAAGGAGCAGGAAGAGTAC 357

Qy 4484 gagaagaaaaatctcgagggcccgaa 4510
Db 358 GATAAAGAAAAACCGATTTCCTCCAAA 384

RESULT 5
US-08-787-335-18
; Sequence 18, Application US/08787335
; Patent No. 5981834
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O BOX 2113
; STREET: FIRST WISCONSIN PLAZA
; CITY: MADISON
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,335
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,797
; FILING DATE:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-88
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1283 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:

; LIBRARY: CKFB15
; CLONE: E9
US-08-787-335-18

Query Match 5.0%; Score 273.4; DB 4; Length 1283;
Best Local Similarity 84.2%; Pred. No. 6.1e-39;
Matches 326; Conservative 0; Mismatches 46; Indels 15; Gaps 1;

Qy 4124 aatacagcttctttctttctatttgattacatggtcattgcaccccttc 4183
Db 13 ACTAAAAATCTTTGCTTTCTATTTGTAAACATGGCTCATACATTTTGTATCCTTC 72

Qy 4184 ttccttttcaacttttactcataagtgctcactagtgcacggtgacacactgtttcg 4243
Db 73 TTCCTTTTCAACTTTTACTTACTGTCTACTAATAATCGGTAGTCACACCGCTCG 132

Qy 4244 gcagcggctcgagctttattcgagacacaaagcaacctcatcagagctcccaaatggct 4303
Db 133 TCAGCGGCTCGACATTATTTCAGACACAAACACCTCATCAGAGCTGCCACAATGGCT 192

Qy 4304 tcaaaatacgaagcagcaagagctgtgaatacgaagccagaatacaaacagccaaag 4363
Db 193 TCAAAATACGAAAAGCACAAGAGTCT-----GAATACAAACAACCAAAA 237

Qy 4364 taccagcaagagtactcaaaacttgagaagcctgaaatgcaaaaggaggaacacacaa 4423
Db 238 TATCAGCAAAAGTACCCAAAACATGAGAAGCCTAAATGCACAAGGAGGAAAAACAAAA 297

Qy 4424 cccgtcaaacagcatgaagagtaccagagtcacacgaatcaaggagcaaaagagtag 4483
Db 298 CCCTGCAACATCATGAAGAGTACCAGAGTCACGCGAATCGAAGGAGCAGGAAGAGTAC 357

Qy 4484 gagaagaaaaatctcgagggcccgaa 4510
Db 358 GATAAAGAAAAACCGATTTCCTCCAAA 384

RESULT 6
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Welens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:

Db 5319 ATATTATTTTAACTTAATAAGGAAATAAAATAAATAAGAA---TACTGAATAA 5375
Qy 2681 cacatctaagcatcaaatttaaccaaatgacacaaatttcattgattagatcaagct 2740
Db 5376 TAAGTCATATTATACATTTTTTAAAAATGTAACATAATTACAAATACGTAACATGTATTA 5435
Qy 2741 tttgagtcttcaaaacataaaaaattacaaaaaaacaaacttaaaatcattatcaat 2800
Db 5436 TAGAATAATAAGAATTTAATATTAAGGATAAATAAATAATTTAAAAATATATTTTTT 5495
Qy 2801 ttgaacacaaagccttgccgaatgctaagagcttaaaaaatggctctttttgtttttt 2860
Db 5496 ATGTCATTTATGTTATATTATATATTAACATGATTA-GTTTTTGAAAAATATTTA 5554
Qy 2861 tgttgcaaacggtggagagaagagggaatgaagattgaccatattttttattatgttt 2920
Db 5555 AATATCATATAATAATAAATAGTTAAATAATAGTATTTTCATACAAAATACTAACT 5614
Qy 2921 taacataataatattaataatattaataacttacttttggtgaatgtgacagtgaggag 2980
Db 5615 TATAAGTATATCATATAATATTATATATATATATATTTATGTGTTTTGATTGGGTGAT 5674
Qy 2981 atacgtaaagtattttaacattatacttttgcaagcagttggctggtctacccaagagt 3040
Db 5675 ATAAGGCTATAAGTATATATGGTTCATTATATATTTATATGTGAATAGATACATAT 5734
Qy 3041 gatcaaa-----gtttgagctgccttcaatgagccaattttgcccataatggataaa 3093
Db 5735 AAGTTAATATATTTATTTGTGTATATGCTGTGTTAAGATAGATATGCATTACAGTTAAG 5794
Qy 3094 ggcaatttggtagttcaactgctcacagaataatgttaaaatgaattaaaataagggtg 3153
Db 5795 GGTATAGTTTTTTTTTTTTTTTTTTGTACATATATAAAAATAGATACTAACAATA 5854
Qy 3154 gcttggtcac 3213
Db 5855 TGCATATTACAAGAATAATATTGCTATAAAATATATATATATATATATATAAAGACAT 5914
Qy 3214 gtaattattatttttaaaataaaattatgttatttagattcttaatttttgggagcatc 3273
Db 5915 -TAAACTATACTAATAGGTAATTAGTTTATTATATCATCTTTTATTATTATAATTTT 5973
Qy 3274 catactataatttcgtaacataatattaaaatagtaataaaagtgaattaacttta 3333
Db 5974 TTTTGTTTTACTTCTTGTCTGTTCTTTTGTATTATAATAACAAATATAAACAATA 6033
Qy 3334 aattacaagcataatattaaattttgaatcaattaatttttattctatttttaatta 3393
Db 6034 TCAGTATTGGAATATAAATAAATTATTCTACATATATGCATATATATATATATA 6093
Qy 3394 atttagtctattttttcaaaataaaatttaaatcaaaataaaataattttt---cctta 3450
Db 6094 TAT 6153
Qy 3451 atgttgaaacaactcatgttatacttcaaaattataagtattattttaccttgatgatt 3510
Db 6154 ATTTTATATATTTTAGTATATCTTTAAAGATATTATTAATTTTATATAGTAGCATAT 6213
Qy 3511 tatttattagtataaattctgattataattatgggtgggatacaatcgctttccactaa 3570
Db 6214 ATGATTTATATTATAACAATATTTTCATTTATATAAATATAGAACATGAACATTTT 6273
Qy 3571 atattttaactatgatt-tataaatttttcaacatcgatatttacttattaatacat 3629
Db 6274 ATTATAACTCATATTGAATATATATTTTATAATGTGATTTTTTACTTATTTTTTAT 6333
Qy 3630 aatttatcataattttatggaaattgagaccaagaacattagagacaaattctataa 3689
Db 6334 ATTATACAATAAAATTTTGAATTCATAAAATGCATGAAATACAAAAAATACAACAA 6393
Qy 3690 caaagacaatttagaaaaaat-gtacttttaggttaattttaagtactcttaaccaaa 3748

Db 6394 AACAAATGATAAAACATTTTTTATTAATATAATATAATATAATATAATATATTTTC 6453
Qy 3749 caaaaattcaaatcaaatgaactaaataagataataacatacggaacatcttactgt 3808
Db 6454 CTGTTATTATTATTCATTTTTTTTTTGATGCTATATATATTATATAATAAATTATA 6513
Qy 3809 aatcttacattcccataattttattatgaaaaataatcttattactcgaactaaatgt 3868
Db 6514 ATATATA---ACAACAAAAATTAATAATAATATACTACTTTTAAATATAACAACAA 6570
Qy 3869 tgtcacaaattattatctaaataaagaaaaacacttaattttataacattttttcata 3928
Db 6571 TACAAGAATATGTATCTATATCAATTATATATATGAATATATAAATATGATAGATA 6630
Qy 3929 atttgaaga 3938
Db 6631 TATAGATAGA 6640

Search completed: September 3, 2000, 03:09:58
Job time: 28544 sec

```

Qy 3200 ttattattcggaatgtaattattatatttttaaaataaaattatgttatttagattcttaat 3259
      ||||| | | | | | | | | | | | | | | | | | | | |
Db 369 TTATATTTCATGTGGTTATAATTATAAAAAGTATATATAGTTTGTAAATTGTAATGATATAA 428

Qy 3260 attttggagcattccatactataaatt---tcgtaacataatattaaaatagtaatat 3315
      | | | | | | | | | | | | | | | | | | | | | | |
Db 429 AATTAGACAGATATAAATTAAATTCAAATATTATATTAAATTTATTATATATGATTAT 488

Qy 3316 aaagtgttaatta-actttaaattacaagcataatattaaatttgaatcaattaattttt 3374
      | | | | | | | | | | | | | | | | | | | | | |
Db 489 TATTGATATTTATATAAATTACATATTGTTATTGTATCATTTAATGATTATATATCATAT 548

Qy 3375 atttctattattttaaattaatttagtctattttttcaaaaataaaatttaaactaaataa 3434
      | | | | | | | | | | | | | | | | | | | | | |
Db 549 CCATATATATATATAAATAATTGAATTATAAATAAATAATTGGCATATTACATTTATAAT 608

Qy 3435 aaataatttttcttaattgttgaacaactcatgtttatacttcaaaattataagttattat 3494
      | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 10

T70055

ID T70055 standard; cDNA; 1283 BP.
AC T70055;
DT 20-AUG-1997 (first entry)
DE Cotton fibre specific cDNA clone E9.
KW cotton; E6; fibre; promoter; transgenic plant; truncated;
KW heterologous gene expression; ds.
OS Gossypium hirsutum strain Coker 312.
PN US5620882-A.
PD 15-APR-1997.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-MAY-1992; US-885970.
PR 19-OCT-1994; US-298829.
PA (CETU) AGRACETUS INC.
PI John M;
DR WPI; 97-235185/21.
PT DNA constructs contg. truncated promoter sequence - for
PT fibre-specific gene expression in cotton plants
PS Example 3; Column 45-48; 48pp; English.
CC T70040-57 are cotton fibre-specific cDNA clones which can be used to
CC obtain genomic clones containing fibre-specific promoters. Claimed DNA
CC constructs comprise a truncated promoter sequence (from one of T70031-38)
CC that promotes preferential gene expression in plant fibre cells, a
CC protein coding sequence not naturally associated with the promoter
CC sequence and a 3' termination sequence. The DNA constructs are useful for
CC expressing foreign genes in fibre-producing plants, esp. to produce
CC transgenic cotton plants with varied cotton fibre characteristics and
CC quality. The present sequence comprises E9 cDNA isolated from clone
CC CKFB15-E9 (CK = Coker; FB15 = 15 day old bolls).
SQ Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

Query Match 5.0%; Score 273.4; DB 1; Length 1283;
Best Local Similarity 84.2%; Pred. No. 2.4e-28;
Matches 326; Conservative 0; Mismatches 46; Indels 15; Gaps 1;

Qy 4124 aatacacgttctttttcttatttgattacacatggtcctatagcattcgtccacctttc 4183
Db 13 ACTAAAAATCTTTGCTTTCTATTTTGAACCATGGCTCATAACTTTTGTCTCCTTC 72
Qy 4184 ttccttttccaacttttactcataagtgtctactagtacccgttagccacactgtttcg 4243
Db 73 TTCTTTTCCAACCTTTTACTCATTACTGTCTCACTAATAATCGGTAGTCACACCGTCTCG 132
Qy 4244 gcagcggctcgacgtttatttcgagacacaaacctcatcagagctcccaactgtgct 4303
Db 133 TCAGCGGCTCGACATTATTTCACAGACAAACAACTCATCAGAGTCCCAATTTGGCT 192
Qy 4304 tcaaaatcacgaaagcacgaagagctgtaatacgaagccagaatacaaacagccaaag 4363
Db 193 TCAAAATACGAAAGCACAAGAGCTCT-----GAATACAAACACCAAAA 237
Qy 4364 tatcacgaagagtactcaaaacttgagaagcctgaaatgcaaaaggaggaacaaacaa 4423
Db 238 TATCAGCAAAAGTACCAAAACATGAGAAGCTTAAATGCACAGAGGAGGAAAAACAAA 297
Qy 4424 ccctgcaaacagcatgaagagtaccagagtcacacgaatcaaaaggagcaaaagagtac 4483
Db 298 CCCTGCAACATCATGAAGAGTACCAGAGTCACCGAATCGAAGGAGCAGGAAGAGTAC 357
Qy 4484 gagaagaaaaatctcgacgggcccga 4510
Db 358 GATAAAGAAAAACCGATTTCCTCCAAA 384

RESULT 11

T43361

ID T43361 standard; cDNA; 974 BP.
AC T43361;

DT 11-MAR-1997 (first entry)
DE Cotton FbLate 2-82A gene cDNA clone A8 (FbLate-1).
KW FbLate; promoter; fibre; transgenic plant; cotton; ds.
OS Gossypium hirsutum.
PN WO9639021-A1.
PD 12-DEC-1996.
PF 06-JUN-1996; U09449.
PR 06-JUN-1995; US-467504.
PA (MONS) MONSANTO CO.
PI John ME;
DR WPI; 97-042726/04.
PT Plant fibre-specific, developmentally regulated FbLate promoter -
PT useful for producing transgenic plants, esp. cotton, with altered
PT fibre properties
PS Claim 8; Page 55-56; 79pp; English.
CC cDNA clones A8 or FbLate-1 (T43361) and A11 or FbLate-2 (T43362)
CC correspond to RNAs prevalent in late development of cotton
CC fibers. They were isolated from a 23-day cotton fibre cDNA
CC library by screening with 24-day fibre cDNA. A8 and A11 are
CC partial clones of the FbLate 2-82A gene. They can be used to
CC identify FbLate promoters (see also T43360) useful for fibre-
CC specific expression of foreign proteins in transgenic plants, esp.
CC cotton fiber.
SQ Sequence 974 BP; 388 A; 161 C; 222 G; 203 T;

Query Match 3.8%; Score 210.6; DB 1; Length 974;
Best Local Similarity 68.1%; Pred. No. 3.6e-20;
Matches 340; Conservative 1; Mismatches 145; Indels 13; Gaps 3;

Qy 4311 acgaaaagcacgaagagctgtaatacgaagccagaatacaaacagccaaagtatcacg 4370
Db 446 AAGAAAAACCGATTTCCTCCCAATGGGAAAGCCTAAGGGCAGGAGAAACATAAAGCCG 505
Qy 4371 aagagtactcaaaacttgagaagcctgaaatgcaaaagg-----aggaaaaacaaac 4424
Db 506 AATATCCGAAATACCTGAGTGAAGGAAAGCTAGATGAGGATAGGAACATAAAGATG 565
Qy 4425 cctgcaaacagcatgaagagtaccagagtcacacgaatcaaaaggacaaaaagagtacg 4484
Db 566 AGTTCCCAAGCATGAAAAAGAGAGGAGAAACCTGAGAAAGCATAGTACCCCTGAG 625
Qy 4485 agaaagaaaatctcgacgggcccgaagatcttcgtagccgtcgagccgggggaattc 4544
Db 626 TGGGTTAAATGCTGAATGGCCGAGTCCATGTTTACTCAGTCTGGCTCGAG-----C 679
Qy 4545 gtcgagccttgaatcatatgacgctggtgcatgtgccatcatcagtaatttcattg 4604
Db 680 ACTAAGCCTTAAGCCATATGACATGGTGCATGTCCATCATCATGAGTAATTTCATGG 739
Qy 4605 tatatcgtaa-tatatagtaataaaaaagatggtgattgggaatgtgtgtgtcattc 4663
Db 740 GATATTGTAATTATATTGTTAATAAAAAAGATGGTGGGAAATGTGTGTGTGATTC 799
Qy 4664 ctccatgcactaatggtgaatctctttgcaatagaaattctaaatggttatagttta 4723
Db 800 ATCCATGTAGCAATGCTGAATCTCTTTGCATGCATAGAGATCTGAATGGTTATAGTTA 859
Qy 4724 tgttatagtgtatgtgttagtgaaktaatttttaattgttatgttaataacacac 4783
Db 860 TGTATATCGTTTGTCTAGTGAATTAATTTGAATGTTATGTAATGTTAATACATC 919
Qy 4784 ttggttgatttatgttat 4802
Db 920 TTGGCTTGATTATGTTTT 938

RESULT 12

T43362

ID T43362 standard; cDNA; 645 BP.
AC T43362;
DT 11-MAR-1997 (first entry)
DE Cotton FbLate 2-82A gene cDNA clone A11 (FbLate-2).

```

Db      61 TTACTCATTACTGTCTCACTAATGATCGGTAGCCACCGTCTCGTCAGCGGGCTGCACAT 120
Qy      4259 ttattcagacacacaagaacctcatcagagctcccacaattggtctcaaaaatcgaaaag 4318
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      121 TTTATCCACACACAAAACAACCTCATCAGAGTGCCACAATTGGCTTCAAATAACGAAAGA 180
Qy      4319 cagcaagagtgttgtaatacgaaaggcagaatacaaacagccaaagtatcacgaagagtac 4378
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      181 CACGAAGAGTCT-----GAATACAAACAGCCAAAAATCATGAAGAGTAC 225
Qy      4379 tcaaaacttggaaagcctgaaatgcaaaaggaggaaaaacaaaaaccctgcaaacagcat 4438
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      226 CCAAAACATGAGAAGCCTGAAATGTACAGGAGGAAAAACAAAACCCTGCAAAACATCAT 285
Qy      4439 gaagagtaccacgagtcacacgaatcaaaggagcaaaaagagtacgayaagaagaaatctc 4498
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      286 GAAGAGTACCACGAGTCACGCCAATCGAAGGAGCACGAAGAGTACGATAAAGAAAAACCC 345
Qy      4499 gacggggcccgaa 4510
        || | | | |
Db      346 GATTTCCTCCAAA 357

```

RESULT 6

T43366
ID T43366 standard; DNA; 519 BP.
AC T43366;
DT 11-MAR-1997 (first entry)
DE Cotton FbLate 2-82A gene cDNA clone All amplified fragment.
KW FbLate; promoter; fibre; transgenic plant; cotton;
KW Gossypium hirsutum; ds.
OS Synthetic.
PN WO9639021-A1.
PD 12-DEC-1996.
PR 06-JUN-1996; U09449.
PR 06-JUN-1995; US-467504.
PA (MONS) MONSANTO CO.
PI John ME;
DR WPI; 97-042726/04.
PT Plant fibre-specific, developmentally regulated FbLate promoter -
PT useful for producing transgenic plants, esp. cotton, with altered
PT fibre properties
PS Example 5; Page 63; 79pp; English.
CC A DNA clone (T43366) was generated by 5'RACE using primers (see
CC also T43364-65) based on FbLate2 clone All (T43362), a partial
CC cDNA clone corresponding to mRNA prevalent in the late development
CC of cotton fibre. The RACE product showed 91.6% similarity at the
CC nucleotide level to the genomic clone, FbLate2-82A (see also
CC T43360). The homology of the RACE product started from nucleotide
CC position 2269 of the FbLate2-28A gene. The ATG initiation codon
CC was identified at position 2315 of the gene.
SQ Sequence 519 BP; 191 A; 127 C; 87 G; 114 T;

```
Qy 4372 agagtactcaaaacttgagaagcctgaatgcaaaaaggagaaaaacaaaaccctgcaa 4431
    | | ||| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db   305 AAATACCCAAAACATGAGAAGCCTAAATGCACAAGGAGGAAAAACAAAACCCTGCCA 364

Qy 4432 acagcatgaagagtaccacgagtcacacgaatcaaaggagcaaaaagagtacgagaaaga 4491
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   365 ACATCATGAAGACTACCACGAGTCACGCGAATCGAAGGAGCACGAAGAGTACGATAAGA 424

Qy 4492 aaatctcgacgggcccgaa 4510
    ||| |||
Db   425 AAAACCCGATTCCCCCAA 443
```

RESULT 7

T13048
ID T13048 standard; cDNA; 1283 BP.
AC T13048;
DT 27-MAY-1996 (first entry)
DE Cotton fibre-specific cDNA clone E9.
KW Cotton; fibre; promoter; transgenic plant; crop improvement; ds.
OS *Gossypium hirsutum* strain Coker 312.
PN US5495070-A.
PD 27-FEB-1996.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-MAY-1992; US-885970.
PA (CETU) AGRACETUS INC.
PI John M;
DI WPI; 96-139095/14.
PT New isolated fibre-specific promoters - used for introducing
PT altered fibre-specific characteristics into plants, partic. cotton.
PS Example 3; Column 45-46; 48pp; English.
CC Cotton cDNA clone E9 (T13048) was isolated from a cDNA library of
CC cotton var. Coker 312 15-day-old boll cells using a subtractive
CC hybridization procedure. The clone hybridises strongly to fiber-
CC RNA and weakly to petal RA. E9 and other fibre-specific cDNA clones
CC (see T13033-47 and T13049-T13050) were used to screen cotton genomic
CC libraries, leading to the isolation of genomic clones (see T13025-32
CC and T13052-53) contg. sequences capable of promoting gene expression
CC in fibre cells.
SQ Sequence 1283 BP; 509 A; 233 C; 251 G; 290 T;

```
Query Match          5.0%; Score 275; DB 1; Length 519;  
Best Local Similarity 85.5%; Pred. No. 1.6e-28;  
Matches 324; Conservative 0; Mismatches 40; Indels 15; Gaps 1;
```

Qy 4132 tcttttctttctatttgattaaccatggctcatagcattcgtcaccctttcttcccttt 4191
|||||
Db 80 TCTTTTCTTTTCATTATGGTTAACCATGGCTCAATACTTTTGTCATCCTTTCTCTCTTT 139

Qy 4192 ccaacttttactcataagtgtctcactagtgacccgtagccacactgtttcgcgagcgc 4251
|||||
Db 140 CCAACTTTTACTCATTACTGTCTCACTAATAATCGGTAGTCACACGCTCTCGTCAGCGGC 199

Qy 4252 tcgacgtttattcgagacacaagaacctcatcacagctcccacaattggcttcaaaaata 4311
|||||
Db 200 TCGACATTTATTCCAGACACAAACCTCATCAGAGCTGCCACAATTGGCTTCAAATA 259

Qy 4312 cgaaaagcacgaagagtctgaatacgaaggcagaatacaaacagccaaagtatcacga 4371
|||||
Db 260 CGAAAAGCACAAAAGAGTCT-----GAATACAACAACCAAAATATCACGA 304

Db 121 TTTTATTTTGGTTTTGGGTTTTGTTGAGTTTTTTAGATAATTATTTTAAATATCTGCAT 180
Qy 1918 aattttctgttatttgaaggatgttcgaatttttttcaaaattgaaacgtttaaga 1977
|||||
Db 181 AATTTTCTGTTATTTGAAAAGGATGTTGGAATTTTTTTCAAAATTGAAACGTTAAGA 240
Qy 1978 atttttactactgcaaaatcagaataagtgaaattgtttttagaagattaaataagt 2037
|||||
Db 241 ATTTTACTACTGCAAAATCAGAATAAGTGAATTTGTTTTTAGAAGATTAAATAAGTT 300
Qy 2038 agtattacgatttttagtttgatttggtggaagtaagtattgttttgaacataattat 2097
|||||
Db 301 AGTATTACGATTTTAGTTTGATTGGTGGAAAGTAATGTATGTTTTGAACATAATTAT 360
Qy 2098 ttgacaataattaagttttctaggaataaacggaaatatcttc-tcttttttgtaaaa 2156
|||||
Db 361 TTGACATAAATAAGTTTCTAGGAAATAACGGAAATATCTCTTTTTTTTTTGTA AAA 420
Qy 2157 ttactaatgcaagaacaaacacggttttggggagcaataatctagcttaagtagtcag 2216
|||||
Db 421 TTACTAATGCAAGAACAACAACGTTTGGGAAGCAATAATCTAGCTTAAAGTAGTCAG 480
Qy 2217 tgtaactctcaaaatctggtcataactcttagctgagttgtgtgctacagtagtaag 2276
|||||
Db 481 TGTAACCTCAAAATCTGGTCATACTTCTAGGCTGAGTTGCTGTCTACAGTAGTAAG 540
Qy 2277 tctatagaacttacctgacaaaacgacatgacgtcagggtcgaatctacaacttttct 2336
|||||
Db 541 TCTATAGAACTTACCTGACAAAACGACATGACGTGAGGGTCGAATCTACAACCTTTCT 600
Qy 2337 ttttcttcaattaacatagttgattcgaagttccgactcataataattattacgatt 2396
|||||
Db 601 TTTTCTCAATTAACATATGGTTGATTCAAGTTCGATCTATAAATAATTATTACGATT 660
Qy 2397 atcaatttcaattacottatatctcattataataataagtcagttcaattcagttt 2456
|||||
Db 661 ATCAATTTCAATTACCTTATATCATCTATTATAAATAAAGTCAGTTCAATTCACTTT 720
Qy 2457 cgaagttcccaaaattttgaattttataaaatttttccctaaaacgaaatagttat 2516
|||||
Db 721 CGAAGTTCCTTAAATTTTGAATTTTATAAATTTATTCCTTAAACCGAAATAGTGAT 780
Qy 2517 atctttcaaaatttaagtttcaatttttcaatccgatttcaatttcaatctttataact 2576
|||||
Db 781 ATCTTTCAAAATTAGTTTCATTTTCAATCCGATTCAATTTCATCTTTTATACTCT 840
Qy 2577 ctattatctataattacataaaatttcaaaatttttgaattattacactttagtcct 2636
|||||
Db 841 CTATGATCTATAATTACATAAATTTCAAACTAATTTTGAATATATACACTTTAGTCCCT 900
Qy 2637 aagttcaaaactataaaattttcacttttagaataatcatttttcaatctaaagcatca 2696
|||||
Db 901 AAGTTCAAAATATAAATTTCACTTTAGAAATTAATCATTTTTCACATCTAAGCATCAA 960
Qy 2697 atttaaccaaatgacacaaatttcatgattagttagatcaagcttttgagttctcaaaa 2756
|||||
Db 961 ATTTAACCAAATGACACAAATTTCTGATTAGTTAGATCAAGCTTTTGAGTCTTCAAAA 1020
Qy 2757 ataaaaatt----acaaaaaaacaaacttaaaatcatttatcaatttgaacaacaaa 2812
|||||
Db 1021 CATAAAATTAACAAAAAACAACAACTTAAATCATTTATCAATTTGAACAACAAA 1080
Qy 2813 gcttggcgaatgctaaagagcttaaaatggcttctttgttttctttgttgcaaacgg 2872
|||||
Db 1081 GCTTGGCCGAATGCTAAGAGCTTAAATGGCTCTTTTGTGTTCTTTTGTGCAACGG 1140
Qy 2873 tggagagaagagggaatgaagattgaccatattttttattatgttttaacataata 2932
|||||
Db 1141 TGGAGAGAGAGGGAATGAAGATTGACCATATTTTATTATGTTTAAACATATAATA 1200
Qy 2933 ttaataatttaatacataattatactttgttgaaatgtgacagtgaggagatacgtaaagta 2992
|||||
Db 1201 TTAATAATTTAATCATAATTATACTTTGGTGAATGTGACAGTGGGAGATACGTAAAGTA 1260

Qy 2993 ttttaacattatacttttgcagcagttggctggtctacccaagagtgatcaaagtttg 3052
|||||
Db 1261 -TATAACATTATACCTTTTGAAGCAGTTGGCTGGTCTATCCAAGAGTGATCAAAGTTTG 1319
Qy 3053 agctgccttcaatgagccaatttttgcataatggataaggcaattgttttagttcaa 3112
|||||
Db 1320 AGCTGCCCTCAATGAGCCAATTTTGGCCATAATGGATAAAGCAATTTGTTAGTTCAA 1379
Qy 3113 ctgctcacagaataatgttaaaatgaaataaaataagtggtgctgacacacacac--- 3169
|||||
Db 1380 CTGCTCACAGAATAATGTTAAATGAAATTAATAAGTGGCTGGTCACACACACACA 1439
Qy 3170 aaaaaaaataatgttggttggttgaattttatattacggaatgtaattattatttta 3229
|||||
Db 1440 AAAAAAACTAATGTTGGTTGGTGAATTTTATATTACGGAATGTAATGTTATATTTA 1499
Qy 3230 aaataaaattatgttttagatttcttaatttttggagcattccataactataatttct 3289
|||||
Db 1500 AAATAAATATGTTATTAGATTCTTAATTTTT-GAGCATTCATACATAATCTCTCGT 1558
Qy 3290 a-acataatataaaatagtaataataaagtgtaattaaactttaaataacagcataat 3348
|||||
Db 1559 ATACATAATATTAATAATAGTAATATAAAGTGTAATTAACTTAAATTACAAGCATAAT 1618
Qy 3349 attaaattttgaatcaattaattttttttctattttttaaatttagtctattttt 3408
|||||
Db 1619 ATTAATTTTGAATCAATTAATTTTATTCTATTATTTAATTAATTAGICTATTTTT 1678
Qy 3409 tcaaaaaataaatttaataataaaaaataatttttccctaatgttgaacaactcatg 3468
|||||
Db 1679 TCAAAATAAAATTTAAATCTAAATAAAATAATTTTCTTAATATT----- 1725
Qy 3469 ttatacttcaaaattataagttattattaccttgatgattttattattagtattataa 3528
|||||
Db 1726 -----ATTAATAAATTTATTCAACATCATATTTACTTATTAAATACATAAA 1773
Qy 3529 ttctgattataattatggtgggatacaatcgcttccactaaatattttaaactatgatt 3588
|||
Db 1774 TTAT----- 1777
Qy 3589 ataaatttatttcaacatgtatatttacttataatacataatttatcataattttatg 3648
|||||
Db 1778 -----AATAATTTATCATAATTTTATG 1799
Qy 3649 gaaattgagaccaagaacattaaagagaacaaattctataacaaagacaatttagaaaa 3708
|||||
Db 1800 GAAATTGAGACCAAGAAACATTAAAGAGAACAAATCTATAACAAAGCAATTTAG-TAAA 1858
Qy 3709 aatgtacttttaggtaatttttaagtaactttaacaaacacaaaaattcaaatcaatga 3768
|||||
Db 1859 AATGTACTTTTAGGTAATTTTAAAGTACTTTAACCAACACAAAAATTCAATCAATGA 1918
Qy 3769 actaaataagataataacatacggaacatcttacttgtaactttacattcccataatt 3828
|||||
Db 1919 ACCAAATAAGATAATATAACATACAGATATCTTACTTGTATTCTTACATGCCGTAATC 1978
Qy 3829 ttattatgaaaaataactcttataactcgaactaaatgttgacacaaattattatctaa 3888
|||||
Db 1979 ATATTATGAAAGTAATATTATATTACCTGAGCCAATGCTCTCAAAATATTATCCAA 2038
Qy 3889 ataaagaa--aaacacttaatttttataacatttttcatattttgaaagattatatt 3946
|||||
Db 2039 AAAAAAATGTTGAATATAATTTTATAACATTTTTCATATTGCAAGATTATATT 2098
Qy 3947 tgtatattttagtataaaattttgacatagattgagcaccttcttaacataatccacca 4006
|||||
Db 2099 TGTATATTACGTAATAATTTTGAACATAGATTGAACACCTTCTTAACATAATCCACCA 2158
Qy 4007 taagtcaagtatgtagatgagaattgttacaacaacgtggggccaaatccaccacaaac 4066
|||||
Db 2159 TAAGTCAAGTATGATAGAGAAATGGTACAAACACGTGGGGCCAAATCCACCAAAAC 2218

Db 2041 AGGTAGTATTACGATTTTGTGTTTGGTGGGAAGTAATGTATGTTTGAACATA 2100
Qy 2093 attatttgacaataaagttttctagggataaaccggaataatctctctttttt 2152
Db 2101 ATTATTGACAAATAAAGTTTCTAGGGAATAACGGAATATCTTCTTTTGT 2160
Qy 2153 aaaattactaatgcaagaacaacaacgttttggggagcaataatctagcttaagtag 2212
Db 2161 AAAATTACTAATGCAAGAACAAACACGTTTGGGGAGCAATAATCTAGCTTAAGTAG 2220
Qy 2213 tcagtgaactctcaaaatctggctataactctagctgagtttctgtgtacagtag 2272
Db 2221 TCAGTGAACCTCTAAAATCTGGCTAATCTTAGCTGAGTTCTGCTGTACAGTAG 2280
Qy 2273 taagtctatagaaacttacctgacaaaacacatgacgtcagggcgaaatctacaactt 2332
Db 2281 TAAGTCTATAGAACTTACCTGACAAAACGACATGACGTGAGGTGCAATCTACAACCTT 2340
Qy 2333 tcctttttctcaattaacatatgttgattcaagttccgactataataatttattacg 2392
Db 2341 TCCTTTTCTCAATTAACATATGGTGATCAAGTCCGATCTATAAATTTATTACG 2400
Qy 2393 atttatcaatttcaattacotttatcatcctattataaataagtcagttcaattcag 2452
Db 2401 ATTTATCAATTTCAATTACCTTATATCATCTATATAAATATAAGTCAGTTCAATTCAG 2460
Qy 2453 ttttcgaagttcccaaaaatttgaattttataaattatccctaaaacggaatag 2512
Db 2461 TTTTCGAAGTTCCAAAAATTTGAATTTTATAAATTTATCCCTAAAACGGAATAG 2520
Qy 2513 ttatatctttcaaatgaatttcatctttcaatccgatttcaatttcatcttttataa 2572
Db 2521 TTATATCTTCAAAATTAAGTTTCAATTTTCAATCCGATTCAATTTCACTCTTTATAA 2580
Qy 2573 ctctctattatctataattacataaaatttcaaatattttgaaattttacactttagt 2632
Db 2581 CTCTCTATTATCTATAATTACATAAAATTTCAAAATTAATTTTGAATTTTACACTTTAGT 2640
Qy 2633 cctaagttcaaaactataaaatttcaatttagaattatcatttttcatctctaagca 2692
Db 2641 CCTAAGTTCAAACTATAAAATTTTCACTTTAGAAATTAATCATTTTTCACATCTAAGCA 2700
Qy 2693 tcaaatttaaccaatgacacaaatttcatgattagtagatcaagcttttgagttctca 2752
Db 2701 TCAAAATTAACCAATGACACAAATTTCAATGATTAGTTAGATCAAGCTTTTGAATCTCA 2760
Qy 2753 aaacataaaaaattacaaaaaaacaaacttaaaatcattatcaatttgaacacaaa 2812
Db 2761 AAACATAAAATTAACAAAAAAACAAACTTAAATCATTTTCAATTTGAACACAAA 2820
Qy 2813 gottggccgaatgctaagagcttaaaatggtcttctttgttttcttttgggcaaacgg 2872
Db 2821 GCTTGGCCGAATGCTAAGAGCTTAAATGAGCTTCTTTGTTCTTTTGTGCAACGG 2880
Qy 2873 tggagagaagagggaatgaagattgacctattttttattatgttttaacataata 2932
Db 2881 TGGAGAGAAGAGGGAATGAAGATTGACCATATTTTTTATTATGTTTAAACATATAATA 2940
Qy 2933 ttaataaatttaatacattatacttttggatgtgacagtgaggagatcgtaaagta 2992
Db 2941 TTAATAATTTAATCATAATTTATCTTTGGTGAATGTGACAGTGGGGAGATACGTAAGTA 3000
Qy 2993 ttttaacattatactttttgcaagcagttggtgtctaccaagagtgatcaagtttg 3052
Db 3001 TTTTAACATTATCTTTTTCGAAGCAGTTGGCTGGTCTACCAAGAGTGATCAAGTTTG 3060
Qy 3053 agctgccttcaatgagccaatttttgcataatggataaaggcaatttggtttagtcaa 3112
Db 3061 AGCTGCCTCAATGAGCCAATTTTGGCCATAATGGATAAAGGCAATTTGTTAGTTCAA 3120
Qy 3113 ctgctcacagaataatgttaaatgaataaaataaggtggcctggtcacacacaaa 3172

Db 3121 CTGCTCACAGAAATGTTAAATGAAATTAATAAGGTGGCCTGGTCACACACAAA 3180
Qy 3173 aaaaaactaatgttggttggtgaattttattattacggaatgtaattatttttttt 3232
Db 3181 AAAAACTAATGTGGTGGTGAATTTATATTACGGAATGTAATATTATTTTAAAA 3240
Qy 3233 taaaattatgttatttagattcttaatttttggagcattccataataatttcgtaac 3292
Db 3241 TAAATATTGTTATTAGATTCTTAATTTTGGAGCATTCACATAAATTTTCGTAAC 3300
Qy 3293 ataattataaaatagtaataaagtgtaattacatttaattacaacataatatta 3352
Db 3301 ATAAATATAAATATAGTAATATAAGTGAATTAACCTTAAATTAACAGCATAATATTA 3360
Qy 3353 aattttgaatcaattaatttttttctattatttttaattatgtctatttttcaa 3412
Db 3361 AATTTTGAATCAATTAATTTTATTCTATTATTATTAAATTAAGTCTATTTTTCAA 3420
Qy 3413 aataaaatttaaatctaaataaaataatttttcttaattgtgaacaactcatgttat 3472
Db 3421 AATAAAATTTAAATCTAAATAAAATAATTTTCTCTAATGTTGAAACACTCATGTAT 3480
Qy 3473 acttcaaaattataagttattattacottgatgtattattatttagtattatttct 3532
Db 3481 ACTTCAAAATATAAGTATTATATTACCTTGATGATTATTATTATAGTATTATTAATCT 3540
Qy 3533 gattataattatggtgggatacaatcgctttccactaaatttttaactatgattttaa 3592
Db 3541 GATTATAATTATGGTGGGATACAACTCGCTTCCACTAAATTTTAACTATGATTATAA 3600
Qy 3593 atttattcaacatcgatatttacttattataacataattttatcataatttttgaaa 3652
Db 3601 ATTTATTCAACATCGTATTTTACTTATTATACATAATTTATCATAATTTTATGGAAA 3660
Qy 3653 ttgagaccaagaacattaagagacaaattctataacaaagacaatttagaaaaaatg 3712
Db 3661 TTGAGACCAAGAAACATTAAGAGAACAAATCTATAACAAAGACAATTTAGAAAAATG 3720
Qy 3713 tacttttaggttaattttaagttactcttaacaaacacaaaaattcaaatcaaatgaacta 3772
Db 3721 TACTTTTAGGTAAATTTAAGTACTCTTAACCAACACAAAAATCAAAATCAATGAACATA 3780
Qy 3773 aataagataataacatacggaacattctactgttaattcattacattccataattttat 3832
Db 3781 AATAAGATAATATAACATACGGAACATCTTACTTGTAATCTTACATTCCATAATTTAT 3840
Qy 3833 tatgaaaaataattattattactcgaaactaaatgtgtgcacaaattattatcctaataa 3892
Db 3841 TATGAAAAATATCTTATATTACTCGAACTAAATGTTGTACAAATATTATCTAAATAA 3900
Qy 3893 agaaaaacacttaatttttataacattttttcatattttgaaagattattttgtata 3952
Db 3901 AGAAAAACACTTAATTTTATAACATTTTTCATATATTGAAAGATTATTTTGTATA 3960
Qy 3953 tttagtataaaattttgacatagattgagcaacttcttaacataatccaccataagtc 4012
Db 3961 TTTAGCTAAAAATATTGACATAGATTGAGCACTTCTTAACATAATCCACCATAAAGTC 4020
Qy 4013 aagtatgtatgagaaattgttacaacaacgtggggccaaatccaccacaaactctc 4072
Db 4021 AAGTATGTAGATGAGAAATGGTACAAACACGTTGGGGCCAAATCCACCAAAACATCTC 4080
Qy 4073 tcattctctctataaaaggcttgctacacatagacaacaatccacacacaaatcacgt 4132
Db 4081 TCATTCTCTCTATAAAAGGCTTGCTACACATAGACAACAATCCACACAAATACAGT 4140
Qy 4133 tcttttctctattttgatttaaccatggtctatgacattcggtcaccctttctctctttc 4192
Db 4141 TCTTTTCTTCTATTGATTACCATGGCTCATAGCAATCGTCACCTTTCTCTCTTTTC 4200
Qy 4193 caacttttactcaatggtctcactagtgacggtagccactgttttcggcagcggt 4252
Db 4201 CAACCTTTACTCATAAGTGTCTCAGTAGGACGGTAGCCACACTGTTTGGGACGGCT 4260

Qy 4561 tatgacgctggtgcattgcccacatcatgcagtaatttcattggtatcgtaaatata 4620
|||||
Db 4561 TATGACGCTGGTGCATGTGCCATCATGCAGTAATTCATGGTATATCGTAATATATA 4620

Qy 4621 gtaataaaaaagatggtgattgggaatgtgtgtgcatctccatgcactaatggt 4680
|||||
Db 4621 GTAATAAAAAAGATGGTGATTGGGAATGTGTGTGCATCTCCATGCCTAATGGT 4680

Qy 4681 gaatctcttgcatacatagaattctaaatggttatgttatgttatagtgatgttg 4740
|||||
Db 4681 GAATCTCTTGCATACATAGAATCTAAATGGTTATAGTTTATGTATAGTGAITGTTG 4740

Qy 4741 tagtgaaktaatttttaattgttatcatgtaacatcacttgcttgattatgtt 4800
|||||
Db 4741 TAGTGAATAATTTTAAATGTTGTATCTAATGTTAACATCACTGGCTTGATTATGTT 4800

Qy 4801 atgttatgtattttactttaatgatattgcattgttgaatttaacattgcttgatca 4860
|||||
Db 4801 ATGTTATGATTTTACTTTAATGATATTGCATGATTGTTAATTAAATTCATTGCTTGATCA 4860

Qy 4861 ttatactcttctactattataaataaattgacactgttttggtaaacctttttacaagtta 4920
|||||
Db 4861 TTATACTCTTCTACTATTATAATTAATGACACTGTTTGGTTAACTTTTACAAGTTA 4920

Qy 4921 agacatgtataaataatgacataataattacaagttttgattcaatgttagctatctta 4980
|||||
Db 4921 AGACATGTATAAATATATGACAATATAATTACAAGTTTGTTCATGTTAGCTATCTTA 4980

Qy 4981 gtatgttattgatcttaattacatttaacaaattccacttaaaattttaataata 5040
|||||
Db 4981 GTATGTTATTGATGATCTTAATTACATTAAACAAATCCACTTAAATTTTAATAATA 5040

Qy 5041 ataacaataattattgtaataataacattaaatgcaacaaaaatgaataaataaaa 5100
|||||
Db 5041 ATAACAATAATTATTGTAATATAATACATTAAATGCAACAAAAATGAAATAAATAAAA 5100

Qy 5101 taaaatagcaataattgttataattgttaataataatgtaccatattcttaactgaa 5160
|||||
Db 5101 TAAAATAGCAATAATTGTTATAATATGTAATATAATGTACCATATTCTTAACGTAA 5160

Qy 5161 atagggtctaactataatccctaaaatttcagtttaaatattttataacctgccatatt 5220
|||||
Db 5161 ATAGGGTCTAACCTATAATCCCTAAAATTTCAGTTTAAATATTTTATACCTGCCATATT 5220

Qy 5221 attagaactcttttaaatatattaaaatttttaattataccaatttaattaaactatta 5280
|||||
Db 5221 ATTAGAACCTCTTTTAAATATATTAAATTTTAAATTATACCAATTTAATTAAACTATTA 5280

Qy 5281 attatcttaactaaaactctaaaattttatttaacctattaattaaattcctaattatctt 5340
|||||
Db 5281 ATTATCTTAACATAAACTCTAAAATTTTATTAACTTAAATTAATTAATTAATTAATCTT 5340

Qy 5341 atctaatttaaaactctaatattcctaatttgatttaaattcttgattatcttaattgt 5400
|||||
Db 5341 ATCTAATTAAAACCTCTAATTATCTTAATTGATTAAATCTTGATTATCTTAATTGTT 5400

Qy 5401 aacctctccaccagctagatgctggaccgaatccggagattacatggcattgaga 5460
|||||
Db 5401 AACCTCTCCACCAGCTAGATGCTGGACCGGAATCCGGAGATTACATCGGCATTGAGA 5460

Qy 5461 tggcctagtagtgatcagggttttctagaggtacccaattgcacctatagtgagtcgt 5518
|||||
Db 5461 TGGCCTAGTAGTGATCAGGGTTTCTAGAGGTACCCAATTGCGCCTATAGTGAGTCGT 5518

RESULT 2

T73865

ID T73865 standard; DNA; 5547 BP.

AC T73865;

DT 26-JAN-1998 (first entry)

DE Cotton fibre promoter clone 4-4(6) construct, pCGN5606 (Version I).

KW promoter; fibre-specific; transcriptional factor; promoter;

KW altered phenotype; colour; melanin; indigo; ss.

OS Gossypium hirsutum cv. coker 130.
FH Key Location/Qualifiers
FT misc_feature 1..65
FT /*tag= a
FT /note= "fragment of pBluescriptII polylinker (as
FT stated in the specification)"
FT misc_feature 57..5494
FT /*tag= b
FT /note= "genomic clone 4-4(6) from lambda phage clone of
FT a cotton Coker 130 genomic library (as stated in
FT the specification)"
FT misc_RNA 65..4163
FT /*tag= c
FT /note= "5' flanking region of the 4-4(6) gene (as
FT stated in the specification)"
FT CDS 4163..4502
FT /*tag= d
FT /note= "corresponds to part of the 4-4(6) ORF (as
FT stated in the specification)"
FT CDS complement (4131..4502)
FT /*tag= i
FT /transl_except= (pos:4170..4172, aa:Xaa)
FT /transl_except= (pos:4182..4184, aa:Xaa)
FT /note= "Xaa = stop codon; No start or stop codons"
FT given, possibly conforms to exon structure.
FT Encodes W21899"
FT misc_feature 4502..4555
FT /*tag= e
FT /note= "synthetic polylinker oligonucleotide containing
FT unique target sites for EcoRI, SmaI, SalI, NheI
FT and BglII"
FT misc_feature 4163..4555
FT /*tag= f
FT /note= "stuffer fragment left in place to facilitate the
FT monitoring of cloning manipulations (as stated in
FT the specification)"
FT 3'UTR 4555..5494
FT /*tag= g
FT /note= "corresponds to the 940 nucleotides downstream of
FT the stop codon and constitutes the 3' flanking
FT region of the 4-4(6) gene (as stated in the
FT specification)"
FT misc_feature 5494..5547
FT /*tag= h
FT /note= "fragment of pBluescriptII polylinker (as stated
FT in the specification)"
PN W09640924-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09897.
PR 07-JUN-1995; US-480178.
PR 01-JUL-1996; ZA-005572.
PA (CALJ) CALGENE INC.
PI McBride K, Pear JR, Perez-Grau L, Stalker DM;
DR WPI; 97-052325/05.
DR P-PSDB; W21899.
PT DNA construct contg. gene of interest controlled by cotton fibre
PT transcriptional factor - used to produce altered phenotype cotton
PT fibre cells expressing genes affecting pigmentation
PS Claim 22; Fig 2A-J; 95pp; English.
CC The present sequence is a 4-4 cotton fibre expression cassette (version
CC I) from promoter construct pCGN5606. The lambda genomic phage clone used
CC to form this construct was designated 4-4(6). DNA constructs containing
CC cotton fibre-specific transcriptional factor promoters are useful to
CC produce cotton fibre cells with altered phenotype, especially altered
CC colour. Genes involved in the production of melanin (e.g. tyrosinase
CC gene and ORF438 encoded protein from Streptomyces antibioticus) and
CC indigo (mono-oxygenase genes possibly in conjunction with a
CC tryptophanase gene) are of interest. The promoters of the invention are
CC reliable and permit expression of a protein selectively in cotton fibre
CC to affect qualities such as fibre strength, length, colour and dyability
CC as required. The construct and methods can also be used for the
CC introduction of other advantageous genes into a cotton plant, e.g. a
CC plant hormone. In particular, fibres from a plant producing coloured

Qy 181 ctgatttacatcctttatatagcgtgaaactacaacaactttagctaaaaaataggata 240
Db 181 CTGATTTCATCCTTTATATAGGCTGAAACTACAACAACCTTAGCTAAAAAATAGGATA 240
Qy 241 acctaatgacaaatcacaatcagatattaaacatgatttttagctaacatttaacaac 300
Db 241 ACCTAATAGCAAAATCACAATCAGATATTAAACCATGATTTTAGCTAACATTTAACAAC 300
Qy 301 tttattgaaactaatttgaattttcatctgctgatatgcccaagatttttaggccactaa 360
Db 301 TTTATTGAAACTAATTGAATATTCATCTGCTGATATGCCCAAGATTTTAGGCCACTAA 360
Qy 361 ccgatttggtggtgaactttaacatgctcatgatttgaactgtttgaacaagttttt 420
Db 361 CCGATTGGTGGTGAACCTTAAACATGTCATGCATTGTGAACGTTTGAACAAGTTT 420
Qy 421 gcattttttactatgaactgtttgattaggttgagtacacactgagcttgaagct 480
Db 421 GCATTATTTACTATATGAACCTGTTGATTAGGTGAGTTACACTGAGCTTGTAGCT 480
Qy 481 cactcaaatttttctaatttctaagtgatcagcaaaacttaggacggcggtacgag 540
Db 481 CACTCAAAATTTTCTAATTTCTAAGTGATCAGCAAACTTAGGACGGCGGTACGAG 540
Qy 541 agctcggattgattttctagtttaataaagaacgatttatgttttaactattatgga 600
Db 541 AGCTCGGATTGATTTCTAGTTAATAAATAGACGATTATGTTTTAAACTATTATGGA 600
Qy 601 ctttttgactatgtaactgtttgggaactttttttgtttttttttttttttttttt 660
Db 601 CTTTGTGACTATGTAACGTGTTGGGACTTTATTTTGTGTTTTATTTGCTTTTTTGG 660
Qy 721 tttagtaattatttttttaactgcaaaattatattgtttttacaactaagtcacagt 720
Db 721 TTTAGTAATTATTATTTTAACTGCAAAATATATGTTTTACAACCTAAGTCACAGTT 720
Qy 781 ttcaaaattccataacttagaatttttcgctgcaaaataaagtaactatttaagtgttt 780
Db 781 TTCAAAATTCATAACTTAGAATTTTCGCTGCAAAATAAAGTAATCATTTAAGTGT 780
Qy 841 tctgtataaaataaataaataaatttaacgagatatttctctaaaaattggaattgat 840
Db 841 TCTGTATAAAATAAATAAATAAATTTAACGAGTATTTCTCTAAAAATTGGAATTGAT 840
Qy 901 ttacaaaattagtagtgcacaaacacatgttttatgttacaggcgatagctctaggc 900
Db 841 TTACAAAATTAGTATGTCAAAACACATGTTTATATGTACAGGCGATATCTCTAGGC 900
Qy 961 aaataacatctaggcggtgttgagtggttacaggcgagtggtcattttgagtaagt 960
Db 901 AAATAACATCTAGCGGGGTTTGGAGTGTTACAGGCGAGTGGGCTCATTTGAGTAAGT 960
Qy 1021 atagttaggccgagtttttagattgcatttcaaggtcaaaagattttgaaactcgatg 1020
Db 961 ATAGTTAGGCGCGAGTTTATGATTGCATATTCAAGGTCAAAAGATTTTGTAACTTCGATG 1020
Qy 1081 ctctgtgataagtatagtagtatttttccaattcttatggcatgtgacattgtggc 1140
Db 1081 CTCGTGTGATAAGTATATAGTATGTTTTTCAATTCTTATGGCATGTGACATTGTGGC 1140
Qy 1141 tattctaattaaattgattgttattattgaaactctgacatctgtttctacaagcatg 1200
Db 1141 TATTCTAATTAAATTGATTGTATTATTGAAATCTGATGCACTGTCTTCAAAAGCATG 1200
Qy 1201 gaactctactgctactgctttctgttaagatacaggttgaagtttaacatgcttactat 1260
Db 1201 GAATCTCATGCTACTGCTTCTGTGTTAAAGATACGATTGCAAGTTTAAACATGCTTACTAT 1260
Qy 1261 ttgtatttgccttgatgctatgtcacattacatgggttgggatgatgtgaagga 1320

Db 1261 TTTGATTTTGTCTTGATGCTATGTCACATACATGGGGTGGGATGATAGGTAGGA 1320
Qy 1321 ggaagttttgacagtttaagtattgacactatctggtgtttaaccacatattgttatg 1380
Db 1321 GGAAGTTTGGACAGTTTAAATGATTGACATATCTGGTGGTTTAAACACATATTGTATG 1380
Qy 1381 gcactcttgactgctgttattggtgctgacggcccatatctgttctggaatttatctgt 1440
Db 1381 GCATCTTGACTGCGGTTATGGTGGCTGACGGCCCATATCTGTTCTGAAATTTATCTGT 1440
Qy 1441 gactctggtggcattgtctacaattattgttgggtgtgttttgatggacgagtcgtggg 1500
Db 1441 GACTCTGGTGGCATTGTCTACAATTTATTTGTTGGTGTGTTTGGATGGACGAGTCGTGG 1500
Qy 1501 gaactctatttgggtgtgttcggagttgggtaggaaatttgcgaataatttgcattgt 1560
Db 1501 GAACCTCTATTGGTGTGTTGCGGAGTTGGGTAGGAAATTTGCAAAAAATTTGCATTGT 1560
Qy 1561 gttttctgaaaaatttgcattaacataatcatgcatctcaattttggtcaattgaac 1620
Db 1561 GTTTTTCTGAAAAATATTGCATTAAACATAATCATGCATTCTCAATTTTGGTCAATTGAAC 1620
Qy 1621 gttataaaattctctatgatactgctgattttattacattatattgtttatgcttga 1680
Db 1621 GTTATAAAATCTCTATGATATCTGATCTGTTTATTACATTATATGTTTATGCTTGA 1680
Qy 1681 gtttaagtcaaacattgagattcatagctcaccatattttaaattcatttcaggcaactg 1740
Db 1681 GTTAAGTCAAACTGAGATTCATAGCTACCCCAATTATTAAATCATTTCAGGCAATCTG 1740
Qy 1741 cagacttaggattggatggcgttcaggagcttgattggtttttctcacatcatatttat 1800
Db 1741 CAGACTTAGGATTGGATGGCGTTTCAGGAGCTTGGATTGTTTCTCACATCATATTTTAT 1800
Qy 1801 taaataattattaattaaaatttttgactgtgtgactaattttcagaattt 1860
Db 1801 TAAATAATTATTAATTAATTTTATGAGACTTTTGGAGTGTCTGACTAATTTTCAGAATTT 1860
Qy 1861 tattttggttttgggttttggtaatttttagataattttttaaattttctgcataat 1920
Db 1861 TATTTTGGTTTGGGTTTTGTGTAATTTTATAGATAATTATTTAAATTTCTGCATAAT 1920
Qy 1921 tttctgttatttgaaggatgttcgaatttttttcaaaattgaaacgtttaagaatt 1980
Db 1921 TTTCTGTATTGTAAGGATGTTGCAATTTTTCAAAATTGAAACGTTTAAAGATT 1980
Qy 1981 tttactactgcaaatcagaataagtgaaattgttttttagaagattaataagttagt 2040
Db 1981 TTTACTACTGCAAAATCAGAATAAGTGAATTTGTTTTAGAAAGATTAAATAAGTTAGT 2040
Qy 2041 attacgatttttagtttgatttgggtggaagtaagtgtttttgaacataattatttg 2100
Db 2041 ATTACGATTTTAGTTGATTGGTGGAAAGTAATGTATGTTTTGAACATAATTATTG 2100
Qy 2101 acaataaattagttttctagggaataaacggaataatctctcttttttgaattac 2160
Db 2101 ACAATAATTAGTTTTCTAGGGAATAAACGGAATATCTCTCTTTTTTGTAAAATTAC 2160
Qy 2161 taatgcaagaacaacaacggttttggggagcaaaataatctagcttaagtagtcagtga 2220
Db 2161 TAATGCAAGACAAACAACGTTTTGGGAGCAAAATATCTAGCTTTAAGTAGTCAGTGTA 2220
Qy 2221 actctcaaatctggtcataactcttagctgagttgtgtgctacagtagtaagtcta 2280
Db 2221 ACTCTCAAAATCTGGTCATAACTCTAGGCTGAGTTTGTCTGCTACAGTAGTAAGTCTA 2280
Qy 2281 tagaaacttacctgacaaaacgacatgacgtcagggtcgaactcacaacttttctttt 2340
Db 2281 TAGAACTTACCTGACAAAACGACATGACGTGAGGTCGAATCTCAACTTTTCTTTTT 2340
Qy 2341 ctccaattacatattggttattcaagttccgatctataaattattacagttattca 2400